Volume 30
Number 380 *Evaluation of variance components from a group of experiments with multiple classifications*
Article 1

May 1951

Evaluation of variance components from a group of experiments with multiple classifications

Walter T. Federer
Iowa State College

Evaluation of Variance Components From A Group of Experiments With Multiple Classifications

By Walter T. Federer

Statistics Department

AGRICULTURAL EXPERIMENT STATION, IOWA STATE COLLEGE

Bureau of Agricultural Economics
United States Department of Agriculture
cooperating

RESEARCH BULLETIN 380 . . . . MAY, 1951 . . . . AMES, IOWA

CONTENTS

Page

Summary 245

Introduction 247

Theoretical results 250

Combining groups of experiments 250

1. Unbiased estimates of the ratio of a variance component to the error mean square or variance 250

2. Some characteristics of the averages of ratios of variance components 255

Average genetic advance due to the selection of the highest yielding individuals rather than a randomly selected set of individuals from a group of m strains 258

3. Formulas for estimating the average genetic progress made by the selection of the highest yielding strain 258

4. Formulas for estimating the average genetic progress made by selection of the k highest yielding strains 260

Expectation of mean squares from a randomized complete block experiment containing all possible crosses of k lines 263

5. Expectation of mean squares 263

6. Variance for general combining ability 267

7. Variance for specific combining ability 268

Unbiased estimates of variance components from the weighted squares of means analysis for the two-way classification, with unequal numbers in the subclasses 269

8. Expectation of the mean squares for unequal numbers in each cell 269

9. Expectation of the mean squares for one or zero individuals in each cell 271

Other methods of analysis for the two-way classification with one or zero individuals in the subclasses 273

10. The method of fitting constants 273

11. Missing plot formula and expectation of mean squares for k values missing in one group of the other classification 277

12. Missing plot formula and expectation of mean squares for k missing values for different entries in each of the two-way classifications 281

13. A simple method for obtaining the method of fitting constants analysis for a two-way classification with incomplete results 283

Experimental results and discussion 285

General description of the data 285

Experimental results on 302 corn experiments for yield 287

14. Analysis of the data 287

15. Correlations among items in the analyses of variance 288

16. Frequency distribution of variance components and of the $\hat{a}_u$ from 302 corn experiments 288

17. Average values of the â's for single crosses, top crosses and double crosses 293

Data from experiments involving more than one tester 295

18. Frequency distribution of yields 295

19. Analyses of variance 295

20. Estimation of the ratios of variance components 300

Data from experiments involving more than one place or location 304

21. Description of the data 304

22. Relative value of combinations of strains, replicates and locations in corn yield trials 305

23. Relative value of combinations of replicates and locations in corn yield tests 307

Literature cited 309

SUMMARY

The findings from the study on the development and application of methods for combining variance components from a group of experiments are:

1. Formulas for the expected value of the ratio of variance components from randomized block and split plot experiments were developed and applied to a set of yield data from experiments on corn. Variances for some of the ratios were developed and applied to the data.

2. The expectation of mean squares was found for two-way classifications of the following nature:

(a) Randomized block experiments involving the $k(k-1)/2$ crosses among k lines.

(b) Randomized block experiments with disproportionate numbers in the subclasses by the weighted squares of means analysis.

(c) Randomized block experiments with a number of missing or incomplete results by the method of fitting constants.

(d) Randomized block experiments with the missing yields calculated and inserted in the table of yields. Two special cases were studied.

3. Although no generalized solution for missing plot values was found, the general solution was indicated in the original thesis (6). Formulas for calculating missing results are presented for two special cases.

4. The average ratio of the variety to the error variance components was highest for experiments involving single crosses, lowest for double crosses and intermediate for top crosses. These results follow expectation.

5. The unweighted mean of the unbiased estimates of a ratio of variance components was found to be satisfactory for estimating the population average ratio. When the number of experiments in the sample was greater than 30 to 40, the standard error of the unweighted mean was low enough for practical purposes. For the 11 tester experiments, the standard error of $\bar{b}$ (arithmetic mean of the unbiased ratios of error to line $\times$ tester variance components) was one-fourth of $\bar{b}$. This is larger than desired for most practical purposes, and either more homogeneous material or a larger number of experiments is required in order to reduce the variance of such an average ratio of variance components as $\bar{b}$. The ratio of the unweighted mean of variance components was more variable than the mean of the ratios.

6. For a fixed number of plots and for a simple randomized complete block design it was found that the greatest genetic advance in the selection of the highest yielding corn double crosses and single crosses was made when four and two replicates were used, respectively; the former type of corn hybrid exhibits relatively low and the latter relatively high genetic variation.

7. The present method for conducting the Iowa Corn Yield Tests is to use three locations or places with six replicates at a place. Considering the relative sizes of the error and the variety (single cross, top cross or double cross) by place variance components, it was found that six locations with two replicates at a place was 11 percent more efficient, and nine locations with one replicate at a place was as efficient as the present method. Also, one location with an infinite number of replicates was about one-half as efficient as the standard.

8. For a fixed number of plots the probability of obtaining the highest yielding corn hybrid was found to be higher if more locations and fewer replicates were used. Also six locations with two replicates at a place, with only two-thirds the number of plots, was found to be better for selecting high yielding hybrids than the present method. The formulas developed in the paper were used to obtain the above results.

Two methods were suggested for combining the variance components from the group of corn experiments. The methods are general and may be applied to similar data from various kinds of experiments.

Evaluation of Variance Components From A Group of Experiments With Multiple Classifications[1]

By Walter T. Federer[2]

In conducting a replicated varietal yield trial, the experimenter usually is interested in obtaining the highest yielding variety or varieties from a group of varieties (or strains). A useful kind of information under such circumstances would be knowledge concerning the optimum allocation of replicates and varieties in any single experiment. Yates (23) and Perotti (12) have given a formula for determining the best combination of replicates and varieties in selecting the highest yielding variety or varieties from a simple randomized complete block design. For the more complex designs, formulas for determining the optimum allocation of varieties, replicates and locations (or years) would be useful.

The present paper deals with the summarization of the data from a group of experiments conducted as a part of the corn improvement program of the Iowa Agricultural Experiment Station and the Bureau of Plant Industry from 1940 through 1946 in Iowa. These data were made available through the kind permission of G. F. Sprague[3] and J. L. Robinson[4]. In the main, the objectives of this study are:

1. To summarize and evaluate the variance components from 302 corn yield trials.
2. To develop or extend the necessary statistical procedures for summarizing and evaluating these data.
3. To determine optimum relative values of the number of replicates and varieties and of replicates, varieties and locations for the selection of higher yielding single crosses, top crosses and double crosses.

The results are divided into two parts, the theoretical and

---

[1] Projects 890 and 163 of the Iowa Agricultural Experiment Station and the Bankhead-Jones Project, Bureau of Agricultural Economics, USDA. Condensed from a thesis submitted to the graduate faculty in partial fulfillment of requirements for the degree Doctor of Philosophy. The complete thesis, doctoral thesis No. 906, is on file in the library of Iowa State College.

For their helpful suggestions and criticisms, the writer extends his thanks to W. G. Cochran and A. M. Mood, formerly of the Statistical Laboratory, Iowa State College, Ames, Iowa; to G. F. Sprague, senior agronomist, USDA, BPISAE, and collaborator, Iowa State College, Ames, Iowa; and to others who reviewed this paper. Dr. Sprague provided for a number of the calculations.

[2] Formerly collaborator, Iowa State College, in cooperation with the Bureau of Agricultural Economics, USDA.

[3] Senior agronomist, USDA, BPISAE, and collaborator with the Iowa Agricultural Experiment Station.

[4] Research professor, Iowa Agricultural Experiment Station.

the experimental. The theoretical results include the development of statistical methods required for the summarization of the data on variance components from corn experiments, the solution of expectation of mean squares for various analyses, and the development of formulas for various combinations of missing plots, for various ratios of variance components, and for estimating average genetic advance. Two types of averages of ratios of variance components were obtained and a third was investigated (6).

The experimental results include the application of the statistical techniques to the corn data to determine the best combination of strains and replicates or of strains, replicates and locations in a single experiment which will, on the average, give the largest probability of obtaining the highest yielding strains of corn, and to compare the averages of ratios from the experiments.

For a given type of design, the error variances from yield trials on corn hybrids would not, in general, be expected to be homogeneous since the fields on which they are planted may have been under different managements. Also, various climatic factors, such as cold weather and heavy rains during germination, etc., have considerable effect upon the experimental error for any particular experiment. The data from the hybrid corn experiments studied in this paper bear out these contentions. Likewise in an experiment involving inbred lines of corn crossed with a group of unrelated testers (or tested at a number of places), estimates of the variance components for lines, $\hat{\sigma}_v^2$, and for testers by lines, $\hat{\sigma}_{vt}^2$, may be expected to vary from experiment to experiment; for example, a sample of lines from an open-pollinated variety might be expected to be more variable than an equal sized sample of lines from a high yielding F_2 population. The variation among lines might be increased considerably, due to specific combining ability, by using an inbred line in place of an open-pollinated variety as the tester parent. For any set of corn strains which differ widely in maturity, the date of planting or time of first killing frost may have considerable effect on the estimate, $\hat{\sigma}_v^2$, of the variety variance component.

For most yield trials grown over a large area, this description appears reasonable. Despite this variability, it still is desirable to combine the data from previous years to determine the most efficient procedure for future experimentation.

An efficient method for summarizing the data on vari-

ance components from a large number of experiments of the nature described above was not immediately clear. Some methods could become extremely difficult due to the variation in number and kind of crosses and number of replicates and due to the fact that the distribution for a variance component is complex (13). There is, however, a fairly simple procedure for obtaining an unbiased estimate of the relative values of variance components from a group of experiments, if it is assumed that the experimenter is interested in the ratio of one of the variance components from the experiment to the error mean square or variance. In addition to the fact that it is possible to obtain an unbiased estimate of the ratio of variance components, the ratios can be combined in a simple manner if all the experiments have the same design and are assumed to give equal information regarding the relative sizes of the variance components. The average ratios of variance components are utilized for designing future experiments of the same nature as those summarized.

The ratios of variance components have an unknown distribution, which is a function of the F distribution and the degrees of freedom, and it is desired to estimate the population parameter, which is also unknown. The individual ratios are random variables from the unknown distribution. In estimating a mean of the ratios, say a_u, where each ratio is an estimate of some parameter, say α_u, the mean, $\bar{a}$, will be subject to two sources of variation:

(1) that due to the sampling errors in estimating the α_u, and
(2) that due to the variation among the α_u.

If the variance among the α_u is equal to or greater than the sampling errors associated with estimating the individual α_u, then the unweighted or arithmetic mean is more efficient than the weighted mean, in the sense that it has a lower variance. Cochran (2) has shown this to be the case for data which are not ratios but are simple estimates. He (2) further stated that in practice the situation was even more favorable for the arithmetic mean than indicated since the weights for a semi-weighted or weighted mean need to be estimated. This estimation results in a corresponding loss in information for these means with no corresponding loss in information on the unweighted mean.

The statistical procedure for combining the analyses from an experiment repeated at several times or locations has been given by Fisher (8) and Yates and Cochran (24). For this case, the analysis of variance must be applicable to each of the experiments, and the individual experiments must be

of equal precision. In the event that the individual experiments are not of equal precision, Smith (15), Cochran (2), and Yates and Cochran (24) have devised procedures for obtaining estimates of the treatment effects and the resulting tests of significance. Yates and Cochran (24) have given the method for combining the results of an experiment conducted at several places when the error degrees of freedom varied from place to place. Yates and Zacopanay (25) have considered a somewhat different situation, that is, the combination of the error variances from a group of cereal experiments in which the different experiments contained different cereals but any one experiment contained only varieties of a certain cereal; in their study the several experiments were combined to obtain an average estimate of the ratio of the sampling to the experimental error variances. The mean ratio of the sampling to the experimental error variances was used to determine the loss in information from sampling relative to complete harvesting of the individual plots in a yield trial.

Methods for estimating variance components have been described for particular types of designs by Yates and Zacopanay (25), Yates and Cochran (24), Winsor and Clarke (20), Satterthwaite (13, 14) and Crump (4, 5); the latter three authors dealt with the more general cases. Crump (5) gave variance components estimates for a two-way classification with disproportionate numbers in the subclasses for the method of "unweighted squares of means" and the method of "expected subclass numbers." Winsor and Clarke (20) and Crump (5) have worked out the expectations of the mean squares for an among-groups and within-groups analysis with disproportionate numbers in the groups. Another special case has been treated in a paper by Sprague and Tatum (17) where estimates of the variance in general and specific combining ability among the crosses within a group of k lines were obtained for each of the lines involved.

THEORETICAL RESULTS

COMBINING GROUPS OF EXPERIMENTS

1. *Unbiased estimates of the ratio of a variance component to the error mean square or variance.*

The fact that the expected ratio is a function of Snedecor's F distribution is used in obtaining an unbiased estimate of the ratio of two variance components. For a randomized complete block experiment with n replicates and m crosses, the ratio of the variety variance component to the error

variance component may be expressed as

$$F = \frac{\dfrac{\hat{\sigma}_e^2 + n\hat{\sigma}_v^2}{\sigma_e^2 + n\sigma_v^2}}{\dfrac{\hat{\sigma}_e^2}{\sigma_e^2}} = \frac{1 + na}{1 + n\alpha} \qquad (1.1)$$

where $a = \hat{\sigma}_v^2/\hat{\sigma}_e^2$ and $\alpha = \sigma_v^2/\sigma_e^2$. If f_e = degrees of freedom for error variance, f_v = degrees of freedom for the variety mean square, and $E\,[\hat{a}] = a = \sigma_v^2/\sigma_e^2$, then

$$E\left[\ a\ \right] = \frac{1}{n}\ E\left[\frac{\hat{\sigma}_e^2 + n\hat{\sigma}_v^2}{\hat{\sigma}_e^2} - 1\right]$$

$$= \hat{a}\,\frac{f_e}{f_e - 2} + \frac{2}{n(f_e - 2)}\ , \qquad (1.2)$$

since the first moment of the F distribution is $f_e/(f_e - 2)$ (19). Therefore, an unbiased estimate of α is

$$\hat{a} = \frac{f_e - 2}{f_e}\left\{a - \frac{2}{n(f_e - 2)}\right\}. \qquad (1.3)$$

In the event that the expectation of the variety mean square is $c\sigma_e^2 + n\sigma_v^2$, the unbiased estimate of the ratio of the variety to error variance components is

$$\hat{a} = \frac{f_e - 2}{f_e}\left\{a - \frac{2c}{n(f_e - 2)}\right\}, \qquad (1.4)$$

where c is a constant. Also, if the variety mean square is of the form $\sigma_e^2 + \dfrac{mn - k - n}{m - 1}\sigma_v^2$, then the unbiased estimate of the ratio of the variety to the error variance components is

$$\hat{a} = \frac{f_e - 2}{f_e}\left\{a - \frac{2(m - 1)}{(mn - k - n)\ (f_e - 2)}\right\}. \qquad (1.5)$$

The ratio of the two variance components for experiments with multiple classifications may be obtained in a similar manner. For example, consider a randomized block experiment (design I, table 1) for which the testers and lines or varieties are randomized within the replicate. The yield equation for the *ighth* plot of such an experiment will be assumed to have the form

$$y_{igh} = \mu + r_i + t_h + v_g + (rv)_{ig} + (rt)_{ih} + (vt)_{gh} + e_{igh} \qquad (1.6)$$

TABLE 1. ANALYSIS AND AVERAGE MEAN SQUARES FOR THREE DESIGNS.

Source of variation	Degrees of freedom	Average values for mean squares
Randomized Complete Block (I)		
Replicates	$n-1$	$\sigma_e^2 + p\sigma_{rv}^2 + m\sigma_{rt}^2 + mp\sigma_r^2$
Varieties	$(m-1) = f_v$	$\sigma_e^2 + p\sigma_{rv}^2 + n\sigma_{vt}^2 + np\sigma_v^2$
Testers	$p-1$	$\sigma_e^2 + m\sigma_{rt}^2 + n\sigma_{vt}^2 + mn\sigma_t^2$
Reps. x testers	$(n-1)(p-1)$	$\sigma_e^2 + m\sigma_{rt}^2$
Var. x testers	$(m-1)(p-1) = f_{vt}$	$\sigma_e^2 + n\sigma_{vt}^2$
Reps. x var.	$(n-1)(m-1) = f_{rv}$	$\sigma_e^2 + p\sigma_{rv}^2$
Error	$(m-1)(n-1)(p-1) = f_e$	σ_e^2
Total	$mnp-1$	Total variance
Split-Plot Design (II)		
Replicates	$(n-1)$	$\sigma_e^2 + m\sigma_{rt}^2 + mp\sigma_r^2$
Testers	$(p-1)$	$\sigma_e^2 + m\sigma_{rt}^2 + n\sigma_{vt}^2 + mn\sigma_t^2$
Error (a)	$(n-1)(p-1)$	$\sigma_e^2 + m\sigma_{rt}^2$
Lines (or var.)	$(m-1) = f_v$	$\sigma_e^2 + n\sigma_{vt}^2 + np\sigma_v^2$
Lines x testers	$(m-1)(p-1) = f_{vt}$	$\sigma_e^2 + n\sigma_{vt}^2$
Error (b)	$p(m-1)(n-1) = f_e$	σ_e^2
Total	$mnp-1$	Total variance
Split-Plot Design (III)		
Places	$p-1$	$\sigma_e^2 + m\sigma_{lr}^2 + n\sigma_{lv}^2 + mn\sigma_l^2$
Reps. within a place	$p(n-1)$	$\sigma_e^2 + m\sigma_{lr}^2$
Varieties	$(m-1) = f_v$	$\sigma_e^2 + n\sigma_{lv}^2 + np\sigma_v^2$
Variety x place	$(m-1)(p-1) = f_{lv}$	$\sigma_e^2 + n\sigma_{lv}^2$
Error	$p(m-1)(n-1) = f_e$	σ_e^2
Total	$mnp-1$	Total variance

where μ is the general mean, r_i the effect common to the ith replicate, t_h the effect common to the hth tester, v_g the effect common to the gth line, $(rv)_{ig}$ the effect common to the gth line in the ith replicate, $(rt)_{ih}$ the effect common to the hth tester in the ith replicate, $(vt)_{gh}$ the effect common to the gth line crossed with the hth tester, and e_{igh} is an effect common to all plots in the experiment; $i = 1, 2, \ldots, n$; $g = 1, 2, \ldots, m$; $h = 1, 2, \ldots, p$.

For a split-plot design composed of n replicates each containing p testers as the whole plot and m varieties or lines crossed with each tester as the split or subplot (design II, table 1) the individual plot yield will be assumed to be expressed by the linear equation

$$y_{igh} = \mu + r_i + t_h + (rt)_{ih} + v_g + (vt)_{gh} + e_{igh}. \tag{1.7}$$

Likewise for a split-plot design consisting of m varieties or lines in each of n replicates, planted at p locations (design III, table 1), the yield of the individual plot will be assumed to have the form

$$y_{igh} = \mu + l_h + (lr)_i + v_g + (lv)_{gh} + e_{igh} \tag{1.8}$$

where $i = (h-1)n + 1, (h-1)n + 2, \ldots, hn$; $g = 1, 2, \ldots, m$; $h = 1, 2, \ldots, p$; and the individual regression coefficients are designated in a similar manner as that for design I. The average values for the various mean squares are given in table 1. The expectations of mean squares for other complex designs may be obtained in a similar manner. The above analyses will be sufficient for the purposes of this study.

For simplicity let f_e be the degrees of freedom for the error variance and f_{vt} the degrees of freedom for the variety $\times$ tester (or place) mean square. For designs I and II (table 1) the unbiased estimate of the ratio of the line or variety by tester variance component to the error variance component is

$$\hat{b} = \frac{f_e - 2}{f_e}\left\{\frac{\hat{\sigma}_{vt}^2}{\hat{\sigma}_e^2} - \frac{2}{n(f_e - 2)}\right\}$$

$$= \frac{f_e - 2}{f_e}\left\{b - \frac{2}{n(f_e - 2)}\right\}. \tag{1.9}$$

A similar formula is obtained for the unbiased estimate of the ratio $\beta = \frac{\sigma_{lv}^2}{\sigma_e^2}$ from design III. Likewise, an unbiased

estimate of $\gamma = \frac{\sigma_{rv}^2}{\sigma_e^2}$ (design I) is

$$\hat{c} = \frac{f_e - 2}{f_e}\left\{\frac{\hat{\sigma}_{rv}^2}{\hat{\sigma}_e^2} - \frac{2}{p(f_e - 2)}\right\}$$

$$= \frac{f_e - 2}{f_e}\left\{c - \frac{2}{p(f_e - 2)}\right\}. \qquad (1.10)$$

The unbiased estimate of $\frac{\sigma_v^2}{\sigma_e^2}$ from designs II or III is

$$\hat{d} = \frac{f_e - 2}{f_e}\left\{\frac{\hat{\sigma}_v^2}{\hat{\sigma}_e^2}\right\} = \frac{d(f_e - 2)}{f_e}. \qquad (1.11)$$

For design I, the unbiased estimate of the ratio of the variety variance component to the error variance component is

$$\hat{e} = \frac{f_e - 2}{f_e}\left\{\frac{\hat{\sigma}_v^2}{\hat{\sigma}_e^2} + \frac{2}{np(f_e - 2)}\right\}$$

$$= \frac{f_e - 2}{f_e}\left\{e + \frac{2}{np(f_e - 2)}\right\}. \qquad (1.12)$$

The unbiased estimate of the ratio $\frac{\sigma_v^2}{\sigma_e^2 + n\sigma_{vt}^2}$ for designs I and II is

$$\hat{f} = \frac{f_{vt} - 2}{f_{vt}}\left\{\frac{\hat{\sigma}_v^2}{\hat{\sigma}_e^2 + n\hat{\sigma}_{vt}^2} - \frac{2}{np(f_{vt} - 2)}\right\}$$

$$= \frac{f_{vt} - 2}{f_{vt}}\left\{f - \frac{2}{np(f_{vt} - 2)}\right\}. \qquad (1.13)$$

A similar formula is obtained for the ratio $\frac{\sigma_v^2}{\sigma_e^2 + n\sigma_{1v}^2}$ in design III.

The ratios of other variance components to the error variance component are obtained in a similar manner.

A simple method for combining the ratios of variance components from a group of experiments is to obtain an arithmetic mean of the ratios, but this mean would be biased. Therefore, in order to obtain an unbiased mean of the ratios,

say a_u, use the unbiased estimates of the ratios, $\hat{a}_u$, and obtain an arithmetic mean of these, thus

$$\bar{a} = \frac{1}{N} \quad \hat{a}_u. \tag{1.14}$$

If

$$\bar{a}' = \frac{1}{N} \sum_{u=1}^{N} a_u, \tag{1.15}$$

then $\bar{a}$ and $\bar{a}'$ will not approach the same mean a since $\bar{a}'$ is biased. Furthermore, the variance of a_u is larger than the variance of $\hat{a}_u$ by a factor $\frac{f_e^2}{(f_e-2)^2}$ [see formula (2.2)] which becomes important for f_e small.

The maximum likelihood estimate of a ratio, say $\hat{a}_u$, of two variance components is the ratio of the two components (6). In order to combine these estimates it is necessary to assume that the $\hat{a}_u$ are all estimates of the same parameter, say a. This assumption does not hold for the corn data and hence this method is not included in the present paper. For a discussion of maximum likelihood solutions of ratios of variance components, the reader is referred to the original thesis (6).

2. *Some characteristics of the averages of ratios of variance components.*

For $a = \frac{c\hat{\sigma}_e^2 + n\hat{\sigma}_v^2}{n\hat{\sigma}_e^2} - \frac{c}{n}$, where the error, $\hat{\sigma}_e^2$, and the variety, $c\hat{\sigma}_e^2 + n\hat{\sigma}_v^2$, mean squares have f_e and f_v degrees of freedom, respectively, the variance of a is

$$V(a) = E[a^2] - [Ea]^2 =$$

$$\frac{2f_e^2\ (f_e + f_v - 2)}{f_v(f_e-4)\ (f_e-2)^2} \left\{ \hat{a} + \frac{c}{n} \right\}^2. \tag{2.1}$$

The variance of $\hat{a}$ is

$$V(\hat{a}) = \frac{(f_e-2)^2}{f_e^2} V(a)$$

$$= \frac{2(f_e + f_v - 2)\ (f_e-2)^2}{f_v(f_e-4)f_e^2} \left\{ \frac{c\hat{\sigma}_e^2 + n\hat{\sigma}_v^2}{n\hat{\sigma}_e^2} \right\}^2. \tag{2.2}$$

Making use of the above results the variance of $\bar{a}$ is (assuming all $a_u = a$)

$$V(\bar{a}) = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{a}_u) =$$

$$\frac{1}{N^2} \sum_{u=1}^{N} \left[\frac{2(f_e + f_v - 2)\ (f_e - 2)^2}{f_v(f_e - 4)f_e^2} \right]_u \left[\frac{c\hat{\sigma}_e^2 + n\hat{\sigma}_v^2}{n\hat{\sigma}_e^2} \right]_u^2 \qquad (2.3)$$

where $\bar{a} = \frac{1}{N} \sum_{u=1}^{N} \hat{a}_u$ and the $\hat{a}_u$ are independent (19).

Since the maximum likelihood solution gives an efficient estimate of the parameter a, any other estimate may be compared with it to determine its relative efficiency. To determine the efficiency of the unbiased estimate $\bar{a}$ (1.14) as compared to $\hat{a}$, the maximum likelihood estimate (6), the variances of the two estimates are compared, thus $\frac{V(\bar{a})}{V(\hat{a})}$. The above variances are computed under the assumption that the $a_u = a$. This assumption does not hold for the data considered in this paper. When the $a_u \neq a$ and where the variance among the a_u is as large or larger than the sampling errors of the individual a_u, the unweighted mean is preferable to the weighted mean. The variance of the unweighted mean, formula (1.14), may be approximated by the formula $\sum_{u=1}^{N} \frac{(\hat{a}_u - \bar{a})^2}{N(N-1)}$. These results have been obtained by Cochran (2) for simple estimates of the mean effect.

As a third average, the individual components may be averaged and the average ratio obtained as the ratio of the averages. Thus, for

$$\bar{\sigma}_e^2 = \frac{1}{N} \sum_{u=1}^{N} (\hat{\sigma}_e^2)_u \text{ and } \bar{\sigma}_v^2 = \frac{1}{N} \sum_{u=1}^{N} (\hat{\sigma}_v^2)_u,$$

$$\bar{a}'' = \frac{\bar{\sigma}_v^2}{\bar{\sigma}_e^2} \quad . \qquad (2.4)$$

This average will have a variance which is approximated by

$$V(\bar{a}'') = \left\{\frac{\bar{\sigma}_v^2}{\bar{\sigma}_e^2}\right\}^2 \left\{\frac{V(\bar{\sigma}_v^2)}{\bar{\sigma}_v^4} + \frac{V(\bar{\sigma}_e^2)}{\bar{\sigma}_e^4} - \frac{2 \text{ cov } \sigma_v^2 \sigma_e^2}{N \bar{\sigma}_v^2 \bar{\sigma}_e^2}\right\}, \quad (2.5)$$

where

$$V(\bar{\sigma}_e^2) = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{\sigma}_e^2)_u \quad (2.6)$$

and

$$V(\bar{\sigma}_v^2) = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{\sigma}_v^2)_u. \quad (2.7)$$

The individual variances $V(\hat{\sigma}_e^2)_u$ and $V(\hat{\sigma}_v^2)_u$ have been given by Crump (4) as

$$V(\hat{\sigma}_v^2) = \frac{2}{n^2}\left\{\frac{(\hat{\sigma}_e^2 + n\hat{\sigma}_v^2)^2}{f_v + 2} + \frac{\hat{\sigma}_e^4}{f_e + 2}\right\} \quad (2.8)$$

and

$$V(\hat{\sigma}_e^2) = \frac{2\hat{\sigma}_e^4}{f_e} \quad (2.9)$$

where the symbols are as defined above. An approximate variance may be obtained by considering the individual variance components as normally distributed independent variates and then proceeding in the same manner as for normal variate theory.

To test the hypothesis that the individual ratios $\hat{a}_u$ may have come from the same population, the following test of significance is proposed:

$$\chi^2 (N - 1 \text{ d.f.}) = \sum_{u=1}^{N} \frac{(\hat{a}_u - \bar{a})^2}{NV(\bar{a})} . \quad (2.10)$$

The above test is an approximation since $\hat{a}_u$ are assumed to be normally distributed or nearly so. If the $\hat{a}_u$ are not normally distributed, then some transformation, such as log $\left[\frac{n+1}{n} + \hat{a}_u\right]$, may be used to make them so.

AVERAGE GENETIC ADVANCE DUE TO THE SELECTION OF THE HIGHEST YIELDING INDIVIDUALS RATHER THAN A RANDOMLY SELECTED SET OF INDIVIDUALS FROM A GROUP OF m STRAINS

3. *Formulas for estimating the average genetic progress made by the selection of the highest yielding strain.*

Since such characters as yield, disease resistance, etc. usually are controlled by a large number of factors, it is often necessary to test large numbers of strains in an experiment. The greater the number of strains tested, the larger will be the probability for obtaining a particularly favorable combination of a large number of factors. Also, for a fixed number of experimental plots, the greater the number of strains tested the less accurate will be the comparisons among the strains. An increase in the block size is usually accompanied by an increase in the error variance. Therefore, if the increase in the error variance is neglected and if it is assumed that the number of plots is fixed, then the following procedure for obtaining the formula given by Perotti (12) may be used to estimate the best combination of strains and replicates for a given ratio of the error and variety variance components.

For m strains or varieties and n replicates, the analysis of variance on an experimental mean basis is

Source of variation	d.f.	Average value of mean squares
Varieties	$m-1$	$\frac{\sigma_e^2}{mn}+\frac{\sigma_v^2}{m}$
Error	$(n-1)(m-1)$	$\frac{\sigma_e^2}{mn}$

The intraclass correlation or the regression of phenotype on genotype is $\frac{\sigma_v^2}{\sigma_v^2+\frac{\sigma_e^2}{n}}$, the standard deviation is $\sigma_v^2+\frac{\sigma_e^2}{n}$ and the average value for selecting the highest yielding variety instead of a randomly selected one from a sample of m from a normal population with mean zero and unit variance is $\bar{x}_m$. Now $\bar{x}_m\sqrt{\sigma_v^2+\frac{\sigma_e^2}{n}}$ would be the average gain in standard deviation units due to the selection of the apparently best individual. Since the correlation between phenotype and

genotype, $\frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_e^2}{n}}$, is not perfect, then the average gain must be weighed by the correlation between observed and actual, thus

$$\frac{\sigma_v^2 \bar{x}_m}{\sigma_v^2 + \frac{\sigma_e^2}{n}} \sqrt{\sigma_v^2 + \frac{\sigma_e^2}{n}} = \frac{\sigma_v^2 \bar{x}_m}{\sqrt{\sigma_v^2 + \frac{\sigma_e^2}{n}}} = \frac{a\sigma_e \bar{x}_m}{\sqrt{a + \frac{1}{n}}} \quad (3.1)$$

The above procedure may be extended to cover more complex situations. The analysis for a split plot design (table 1), on an experimental mean basis, is

Source of variation	d.f.	Average value of mean square
Varieties	$(m-1)$	$\sigma_e^2/pmn + \sigma_{vt}^2/pm + \sigma_v^2/m$
Varieties x testers	$(p-1)(m-1)$	$\sigma_e^2/pmn + \sigma_{vt}^2/pm$
Error	$p(m-1)(n-1)$	σ_e^2/pmn

In place of testers, years or locations may be used in a similar analysis. The intraclass correlation is $\frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_{vt}^2}{p} + \frac{\sigma_e^2}{np}}$ and the standard deviation is $\sqrt{\sigma_v^2 + \sigma_{vt}^2/p + \sigma_e^2/np}$. Thus the average gain or genetic advance due to the selection of the highest yielding instead of a randomly selected variety from a sample of m lines is

$$\frac{\sigma_v^2 \bar{x}_m}{\sqrt{\sigma_v^2 + \frac{\sigma_{vt}^2}{p} + \frac{\sigma_e^2}{np}}} = \frac{\delta \sigma_e \bar{x}_m}{\sqrt{\delta + \frac{\beta}{p} + \frac{1}{np}}} . \quad (3.2)$$

Likewise for a design in which a group of m lines are crossed with p testers and the resulting crosses grown in n completely randomized blocks, the following pertinent part of the analysis (table 1) is

Source of variation	d.f.	Average value of mean square
Varieties (or lines)	$(m-1)$	$\frac{\sigma_e^2}{mnp}+\frac{\sigma_{rv}^2}{mn}+\frac{\sigma_{vt}^2}{mp}+\frac{\sigma_v^2}{m}$
Varieties x replicates	$(n-1)(m-1)$	$\sigma_e^2/mnp+\sigma_{rv}^2/mn$
Varieties x testers	$(p-1)(m-1)$	$\sigma_e^2/mnp+\sigma_{vt}^2/mp$
Varieties x replicates x testers	$(p-1)(m-1)(n-1)$	σ_e^2/mnp

In this case the average progress made by selecting the highest yielding variety is

$$\frac{\sigma_v^2\bar{x}_m}{\sqrt{\sigma_v^2+\frac{\sigma_{rv}^2}{n}+\frac{\sigma_{vt}^2}{p}+\frac{\sigma_e^2}{np}}}=\frac{\delta\,\sigma_e\bar{x}_m}{\sqrt{\delta+\frac{\gamma}{n}+\frac{\beta}{p}+\frac{1}{np}}}, \qquad (3.3)$$

which reduces to the preceding formula if $\sigma_{rv}^2=0$.

The above process may be extended to other situations by the method outlined above.

4. *Formulas for estimating the average genetic progress made by selection of the* k *highest yielding strains.*

In section 3, formulas were given for obtaining the average genetic advance due to the selection of the highest yielding rather than a randomly selected line from a sample of m lines. The expected value for the largest deviate may be obtained from the table of ranges given by Tippett (18). These tables were reproduced by Karl Pearson (11) in table XXII and in part by Snedecor (16). If it is desired to obtain the expected value for the k highest yielding strains from a sample of size m, the expected values for the highest, the next highest, etc., down to the kth highest yielding individuals are averaged. For example, the expected value for the two highest yielding strains from a sample of size 50 is $\frac{(2.25+1.85)\sigma}{2}=2.05\sigma$ where σ is the standard deviation and the values 2.25 and 1.85 were obtained from table XX of Fisher and Yates (9). Unfortunately the expected values for the second largest, the third largest, etc., have not been tabulated for samples of m greater than 50.

Since the formula for the expected value of the kth largest deviate (9, 19) involved tedious numerical integration, some simpler solution is desired. The values for the expected value of the range have been tabled by Tippett (18) for values of m ranging from 2 to 1,000. The expected value of the largest deviate is one-half of the range. These values are given in table 2 for $m = 41$ to 100 and in table 3 for $m = 101$ to 200.

The expected value of the largest deviate from a sample of size m is given as (9, 19),

$$E(x_{m,1}) = m \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x_1} f(x)dx \right]^{m-1} f(x_1)dx_1$$

$$= m \int xF^{m-1}\,dF. \tag{4.1}$$

Formula (4.1) may be written as

$$E(x_{m,1}) = m \int xF^{m-2}\,dF - m \int xF^{m-2}(1-F)dF$$

$$= \frac{m}{m-1} E(x_{m-1,1}) - \frac{1}{m-1} E(x_{m,2}). \tag{4.2}$$

Therefore, the expected value for the second largest member from a sample of size m may be obtained from the largest members of samples of size m and $m-1$, thus

$$E(x_{m,2}) = mE(x_{m-1,1}) - (m-1)E(x_{m,1}). \tag{4.3}$$

The recursion formulas for obtaining the third, fourth and fifth largest deviates, respectively, were found by the same device to be

$$E(x_{m,3}) = \frac{1}{2}\left\{mE(x_{m-1,2}) - (m-2)E(x_{m,2})\right\}, \tag{4.4}$$

$$E(x_{m,4}) = \frac{1}{3}\left\{mE(x_{m-1,3}) - (m-3)E(x_{m,3})\right\}, \tag{4.5}$$

$$E(x_{m,5}) = \frac{1}{4}\left\{mE(x_{m-1,4}) - (m-4)E(x_{m,4})\right\}. \tag{4.6}$$

Recursion formulas for other ordered deviates may be set up in a similar manner. However, due to the fact that the range tabulated by Tippett (18) was correct to four decimal places only, it is not possible to obtain the expected values of the third largest deviates for samples of 100 or less correct to more than one decimal place. If the expected value

TABLE 2. AVERAGE VALUES FOR THE FIRST, SECOND, THIRD, FOURTH AND FIFTH LARGEST DEVIATES FROM SAMPLE SIZES OF 41 TO 100 FROM A NORMAL DISTRIBUTION WITH MEAN ZERO AND UNIT STANDARD DEVIATION.

Sample size	Value of deviate					Sample size	Value of deviate				
	1	2	3	4	5		1	2	3	4	5
41	2.171	1.765	1.530	1.36	1.22	71	2.383	2.007	1.793	1.63	1.52
42	2.180	1.776	1.542	1.37	1.23	72	2.388	2.013	1.799	1.64	1.53
43	2.190	1.787	1.554	1.38	1.25	73	2.393	2.019	1.806	1.65	1.53
44	2.199	1.797	1.567	1.40	1.26	74	2.398	2.025	1.813	1.65	1.53
45	2.208	1.807	1.580	1.41	1.27	75	2.403	2.031	1.819	1.66	1.54
46	2.216	1.817	1.593	1.42	1.28	76	2.408	2.036	1.825	1.66	1.54
47	2.225	1.827	1.603	1.43	1.30	77	2.413	2.042	1.832	1.67	1.54
48	2.233	1.837	1.614	1.44	1.31	78	2.417	2.047	1.838	1.68	1.55
49	2.241	1.847	1.624	1.45	1.32	79	2.422	2.052	1.844	1.69	1.55
50	2.249	1.856	1.634	1.46	1.33	80	2.427	2.057	1.850	1.69	1.56
51	2.257	1.864	1.643	1.48	1.34	81	2.431	2.063	1.855	1.70	1.56
52	2.264	1.873	1.654	1.49	1.36	82	2.436	2.068	1.861	1.71	1.56
53	2.272	1.881	1.662	1.50	1.37	83	2.440	2.073	1.867	1.71	1.57
54	2.279	1.889	1.669	1.51	1.38	84	2.445	2.077	1.872	1.72	1.57
55	2.286	1.897	1.677	1.52	1.39	85	2.449	2.082	1.877	1.73	1.58
56	2.293	1.905	1.685	1.53	1.40	86	2.453	2.087	1.883	1.73	1.58
57	2.300	1.912	1.692	1.54	1.42	87	2.457	2.092	1.888	1.74	1.59
58	2.306	1.920	1.700	1.55	1.43	88	2.462	2.096	1.893	1.75	1.60
59	2.313	1.927	1.708	1.56	1.44	89	2.466	2.101	1.898	1.75	1.61
60	2.319	1.935	1.715	1.56	1.45	90	2.470	2.105	1.902	1.76	1.61
61	2.326	1.942	1.722	1.57	1.46	91	2.474	2.110	1.907	1.77	1.62
62	2.332	1.949	1.729	1.57	1.47	92	2.478	2.114	1.912	1.77	1.63
63	2.338	1.956	1.737	1.58	1.48	93	2.482	2.119	1.916	1.78	1.64
64	2.344	1.963	1.744	1.59	1.49	94	2.485	2.123	1.920	1.78	1.65
65	2.350	1.969	1.751	1.60	1.49	95	2.489	2.127	1.925	1.79	1.66
66	2.355	1.976	1.758	1.60	1.50	96	2.493	2.131	1.929	1.80	1.67
67	2.361	1.983	1.765	1.61	1.51	97	2.497	2.136	1.933	1.80	1.68
68	2.367	1.989	1.772	1.61	1.51	98	2.500	2.140	1.937	1.81	1.69
69	2.372	1.995	1.779	1.62	1.52	99	2.504	2.144	1.941	1.81	1.70
70	2.377	2.001	1.786	1.63	1.52	100	2.508	2.148	1.945	1.81	1.70

of every fifth or sixth largest deviate were obtained correctly to several significant figures, say 8 to 10, by numerical integration, then it would be possible to use the above recursion formula and obtain the values for the ordered deviates in between by successive approximations.

Because data on expected values of deviates are not available to several significant figures, the recursion formulas are not satisfactory in themselves. This fact is brought out by a study of the values given in tables 3 and 4 of the thesis (6). However, by computing a fifth degree regression on the values given by Tippett (18) for the largest deviate, equal one-half the expected range, the values were computed to nine significant figures. Recursion formula (4.3) was then applied to these values and the expected values for the second largest deviates obtained. A fifth degree regression was fitted to the latter values to obtain the second largest deviates correct to seven decimals. Recursion formula (4.4) was applied to obtain the third largest deviates. This process was continued for the fourth and fifth largest deviates. These values are given in tables 2 and 3 to four significant figures for the first, second and third largest deviates and to three significant figures for the fourth and fifth.

These data are suitable for most practical conditions encountered. The agreement of the values in tables 2 and 3 with those given by Fisher and Yates (9) was good, but it is suggested that the accuracy of this method be tested further before many more additional expected values are computed.

A rough approximation of the expected value of the largest k individuals (k less than $m/2$) may be obtained from the next to the last column of Pearson's (11) table II. This approximation will approach the true value as the sample size m increases. If $k > m/2$, then Pearson's (10) table XI should be used.

For the value $\bar{x}_m$ in formulas (3.1), (3.2) and (3.3), an average of the values in table 2 may be used to obtain the average genetic progress made by selecting the highest two, three, four or five strains from samples of size 41 to 100. Likewise, the values in table 3 may be used to obtain the average advance due to the selection of the two, three or four highest yielding strains from samples of size 101 to 200.

EXPECTATION OF MEAN SQUARES FROM A RANDOMIZED COMPLETE BLOCK EXPERIMENT CONTAINING ALL POSSIBLE CROSSES OF k LINES

5. *Expectation of mean squares.*

A number of breeding experiments are conducted using all possible crosses among a group of k lines. In some cases

TABLE 3. AVERAGE VALUES FOR THE FIRST, SECOND, THIRD AND FOURTH LARGEST DEVIATES FROM SAMPLE SIZES OF 101 TO 200 FROM A NORMAL DISTRIBUTION WITH MEAN ZERO AND UNIT STANDARD DEVIATION.

Sample size	Value of deviate				Sample size	Value of deviate			
	1	2	3	4		1	2	3	4
101	2.511	2.152	1.949	1.82	151	2.652	2.309	2.119	1.98
102	2.515	2.156	1.952	1.82	152	2.654	2.311	2.122	1.99
103	2.518	2.160	1.956	1.83	153	2.656	2.314	2.124	1.99
104	2.522	2.164	1.960	1.84	154	2.658	2.316	2.127	1.99
105	2.525	2.168	1.964	1.84	155	2.660	2.319	2.129	2.00
106	2.528	2.172	1.968	1.85	156	2.663	2.321	2.132	2.00
107	2.532	2.175	1.972	1.85	157	2.665	2.324	2.134	2.00
108	2.535	2.179	1.977	1.85	158	2.667	2.326	2.137	2.01
109	2.538	2.183	1.981	1.86	159	2.669	2.328	2.139	2.01
110	2.541	2.186	1.985	1.86	160	2.671	2.331	2.142	2.01
111	2.545	2.190	1.989	1.86	161	2.673	2.333	2.144	2.01
112	2.548	2.193	1.993	1.86	162	2.675	2.335	2.147	2.02
113	2.551	2.197	1.997	1.87	163	2.678	2.338	2.149	2.02
114	2.554	2.201	2.001	1.87	164	2.680	2.340	2.152	2.02
115	2.557	2.204	2.005	1.87	165	2.682	2.342	2.154	2.02
116	2.560	2.207	2.009	1.87	166	2.684	2.345	2.156	2.03
117	2.563	2.210	2.013	1.88	167	2.686	2.347	2.159	2.03
118	2.566	2.214	2.016	1.88	168	2.688	2.349	2.161	2.03
119	2.569	2.217	2.020	1.88	169	2.690	2.351	2.163	2.03
120	2.572	2.220	2.023	1.89	170	2.692	2.354	2.165	2.04
121	2.575	2.224	2.027	1.89	171	2.694	2.356	2.168	2.04
122	2.578	2.227	2.030	1.89	172	2.696	2.358	2.170	2.04
123	2.581	2.230	2.034	1.89	173	2.698	2.360	2.172	2.04
124	2.584	2.233	2.037	1.90	174	2.700	2.362	2.174	2.04
125	2.586	2.236	2.040	1.90	175	2.701	2.364	2.177	2.05
126	2.589	2.239	2.044	1.90	176	2.703	2.366	2.179	2.05
127	2.592	2.242	2.047	1.91	177	2.705	2.369	2.181	2.05
128	2.595	2.246	2.050	1.91	178	2.707	2.371	2.183	2.05
129	2.597	2.249	2.054	1.91	179	2.709	2.373	2.186	2.05
130	2.600	2.252	2.057	1.92	180	2.711	2.375	2.188	2.05
131	2.603	2.255	2.060	1.92	181	2.713	2.377	2.190	2.05
132	2.605	2.257	2.063	1.92	182	2.715	2.379	2.193	2.05
133	2.608	2.260	2.066	1.93	183	2.716	2.381	2.195	2.05
134	2.610	2.263	2.070	1.93	184	2.718	2.383	2.197	2.06
135	2.613	2.266	2.073	1.93	185	2.720	2.385	2.200	2.06
136	2.616	2.269	2.076	1.94	186	2.722	2.387	2.202	2.06
137	2.618	2.272	2.079	1.94	187	2.724	2.389	2.204	2.06
138	2.621	2.275	2.082	1.94	188	2.725	2.391	2.206	2.06
139	2.623	2.277	2.085	1.95	189	2.727	2.393	2.208	2.06
140	2.626	2.280	2.088	1.95	190	2.729	2.395	2.210	2.06
141	2.628	2.283	2.091	1.95	191	2.731	2.397	2.212	2.07
142	2.630	2.286	2.094	1.95	192	2.732	2.399	2.214	2.07
143	2.633	2.288	2.097	1.96	193	2.734	2.401	2.216	2.07
144	2.635	2.291	2.100	1.96	194	2.736	2.402	2.219	2.07
145	2.638	2.293	2.102	1.96	195	2.738	2.404	2.221	2.07
146	2.640	2.296	2.105	1.97	196	2.739	2.406	2.223	2.07
147	2.642	2.299	2.108	1.97	197	2.741	2.408	2.225	2.07
148	2.645	2.301	2.111	1.97	198	2.743	2.410	2.227	2.07
149	2.647	2.304	2.114	1.98	199	2.744	2.411	2.229	2.07
150	2.649	2.306	2.116	1.98	200	2.746	2.413	2.231	2.07

the reciprocal crosses are identical and hence are not included in the experiment, for example, yield trials involving single-cross or double-cross corn hybrids. Such an experiment would have $k(k-1)/2$ crosses, and if the design were a randomized complete block, the yield of an individual

plot will be assumed to be represented by the linear equation

$$y_{ij} = \mu + r_i + v_j + e_{ij}, \tag{5.1}$$

where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k(k-1)/2$; μ is the general mean, r_i is an effect common to the ith replicate, v_j is an effect common to the jth variety, and e_{ij} is an effect common to all plots. If it is desired to consider the general or over-all effect of k lines, then the yield of the ighth plot is expressed as

$$y_{igh} = \mu + r_i + l_g + l_h + (lt)_{gh} + e_{igh}, \tag{5.2}$$

where $g \neq h = 1, 2, \ldots, k$; the double subscript gh runs from $g < h = 2, 3, \ldots, k$; l_g is the effect common to the gth line, l_h is the effect common to the hth line, $(lt)_{gh}$ is the effect common to the cross of the gth by the hth line, and e_{igh} is an effect common to all plots.

The analysis of variance, considering the line effects, for such an experiment is:

Source of variation	Degrees of freedom	Sum of squares
Replicates	$n-1$	$\sum_{i=1}^{n} \frac{2Y_{i\cdot\cdot}^2}{k(k-1)} - \frac{2Y_{\cdot\cdot\cdot}^2}{nk(k-1)}$
Varieties	$(m-1) = \frac{k(k-1)}{2} - 1$	$\sum_{g<h=2}^{k} \frac{Y_{\cdot gh}^2}{n} - \frac{2Y_{\cdot\cdot\cdot}^2}{nk(k-1)}$
Among lines	$k-1$	$\sum_{g=1}^{k} \frac{4(\frac{k}{2}Y_{\cdot g\cdot} - Y_{\cdot\cdot\cdot})^2}{nk^2(k-2)}$
Within lines	$\frac{k(k-3)}{2}$	Subtraction
Error	$(n-1)(m-1)$	$\sum_{i=1}^{n}\sum_{g<h=2}^{k} y_{igh}^2 - \sum_{i=1}^{n} \frac{2Y_{i\cdot\cdot}^2}{k(k-1)} - \sum_{g<h=2}^{k} \frac{Y_{\cdot gh}^2}{n} + \frac{2Y_{\cdot\cdot\cdot}^2}{nk(k-1)}$
Total	$\frac{nk(k-1)}{2} - 1$	$\sum_{i=1}^{n}\sum_{g<h=2}^{k} y_{igh}^2 - \frac{2Y_{\cdot\cdot\cdot}^2}{nk(k-1)}$

The expectation of the replicate, variety and error mean squares are $\sigma_e^2 + \frac{k(k-1)}{2}\sigma_r^2$, $\sigma_e^2 + n\sigma_v^2$, and σ_e^2, respectively, which follow from the randomized complete block analysis. Using the regression equation (5.2) for y_{igh}, the various totals are:

$$Y._{gh} = n\mu + \sum_{i=1}^{n} r_i + n\left\{ l_g + l_h + (lt)_{gh} \right\} + \sum_{i=1}^{n} e_{igh} \quad (5.3)$$

$$Y._{g.} = n(k-1)\mu + (k-1)\sum_{i=1}^{n} r_i + n(k-1)l_g + n\sum_{\substack{h=1 \\ \neq g}}^{k}\{l_h + (lt)_{gh}\} + \sum_{i=1}^{n}\sum_{\substack{h=1 \\ \neq g}}^{k} e_{igh} \quad (5.4)$$

$$Y... = \frac{k(k-1)}{2}\left\{ n\mu + \sum_{i=1}^{n} r_i \right\} + n(k-1)\sum_{g=1}^{k} l_g + n\sum_{g<h=2}^{k}(lt)_{gh} + \sum_{i=1}^{n}\sum_{g<h=2}^{k} e_{igh}. \quad (5.5)$$

The expectation of the sum of squares among the $k(k-1)/2$ crosses is

$$\left\{ \frac{k(k-1)}{2} - 1 \right\}\left\{ \sigma_e^2 + n\sigma_{lt}^2 \right\} + n(k-1)(k-2)\sigma_l^2. \quad (5.6)$$

The preceding expectations may also be obtained in the following sum of squares

$$\left[E \sum_{g<h=2}^{k} 4\,\frac{\left\{ \frac{k(k-1)}{2} Y._{gh} - Y... \right\}^2}{nk^2(k-1)^2} \right] \quad (5.7)$$

with $\frac{k(k-1)}{2} - 1 = \frac{(k+1)(k-2)}{2}$ degrees of freedom.

The expectation of the line sum of squares is

$$E\left[\sum_{j=1}^{k} \frac{4\left[\frac{k}{2} Y._{g.} - Y... \right]^2}{nk^2(k-2)} \right] = (k-1)\left\{ \sigma_e^2 + n\sigma_{lt}^2 + n(k-2)\sigma_l^2 \right\} \quad (5.8)$$

with $k-1$ degrees of freedom. The divisor of equation (5.8) is $nk^2(k-2)/4$ since the $n(k-2)$ quantities l_g are

multiplied by $k/2$. The within line sum of squares is obtained by subtraction and has the expectation

$$\frac{k(k-3)}{2}\left\{\sigma_e^2 + n\sigma_{lt}^2\right\} \tag{5.9}$$

with $k(k-3)/2$ degrees of freedom.

For a paper written by Sprague and Tatum (17) W. G. Cochran had derived formulas for specific and general combining ability from an experiment such as that described above. The derivation of these estimates as indicated by Cochran follows.

6. *Variance for general combining ability.*

The yield for any particular plot is represented by the equation for y_{lgh} as given previously. Now if the k lines are considered as the population it may be assumed that

$$\sum_{\substack{h=1\\ \neq g}}^{k} (lt)_{gh} = 0 \text{ and } \sum_{\substack{h=1\\ \neq g}}^{k} l_h = -l_g .$$

Making use of these assumptions, the line and grand totals become

$$Y_{.g.} = (k-1)\,(n\mu + \sum_{i=1}^{n} r_i) + n(k-2)l_g + \sum_{i=1}^{n}\sum_{\substack{h=1\\ \neq g}}^{k} e_{igh} \tag{6.1}$$

$$Y... = \frac{k(k-1)}{2}\,(n\mu + \sum_{i=1}^{n} r_i) + \sum_{i=1}^{n}\sum_{g<h=2}^{k} e_{igh} = \sum_{g=1}^{k} \frac{Y_{.g.}}{2} . \tag{6.2}$$

The difference, $\left(\frac{k}{2}Y_{..g.} - Y...\right)$, squared for any particular line g has the expectation

$$\frac{n^2(k-2)^2}{4}k^2\sigma_l^2 + \frac{nk(k-1)\,(k-2)}{4}\sigma_e^2 . \tag{6.3}$$

Therefore,

$$\hat{\sigma}_{l_g}^2 = \frac{(k-1)}{k(k-2)} \left\{ \frac{4(\frac{k}{2} Y_{.g.} - Y_{...})^2}{n^2 k(k-1)(k-2)} - \frac{\hat{\sigma}_e^2}{n} \right\} \tag{6.4}$$

which is a form of the formula obtained by Cochran.

7. *Variance for specific combining ability.*

Using the equation for y_{lgh}, the sum of squares for the effect of a particular line g crossed on the other $(k-1)$ lines has the expectation

$$E\left[\sum_{h=1 \neq g}^{k} \left\{ \frac{(k-2)Y_{.gh} - Y_{.g.} - Y_{..h} + \frac{2Y_{...}}{(k-1)}}{n(k-2)(k-3)} \right\}^2 \right] = (k-2)\left\{ \sigma_e^2 + n\sigma_{lt}^2 \right\} \tag{7.1}$$

with $(k-2)$ degrees of freedom. If the $\sum_{h=1 \neq g}^{k} (lt)_{gh}$ is assumed to be zero, then the expectation of the above sum of squares was incorrectly given by Sprague and Tatum (17) as

$$n(k-2)\left\{ \frac{\sigma_e^2}{n} + \frac{(k-3)}{(k-2)} \sigma_{lt}^2 \right\} . \tag{7.2}$$

The correct expectation has been ascertained[5] to be

$$n(k-2)\left\{ \frac{\sigma_e^2}{n} + \frac{(k-1)}{(k-3)} \sigma_{lt}^2 \right\} . \tag{7.3}$$

Sprague and Tatum (17) illustrate the use of these formulas with an example on corn hybrids involving $k = 10$ inbred lines or 45 single crosses of corn.

Considering the $(k-1)$ lines crossed with the line g as a sample of lines, the variance for the mean of the gth line is

$$E\left[\frac{Y_{.g.}}{n(k-1)} \right]^2 - \left[E \frac{Y_{.g.}}{n(k-1)} \right]^2 = \frac{\sigma_e^2}{n(k-1)} + \frac{\sigma_{lt}^2}{(k-1)} \tag{7.4}$$

since

$$\sum_{i=1}^{n} r_i = 0 \quad \text{and} \quad \sum_{h=1 \neq g}^{k} l_h = -l_g .$$

[5] By written correspondence with W. G. Cochran.

UNBIASED ESTIMATES OF VARIANCE COMPONENTS FROM THE WEIGHTED SQUARES OF MEANS ANALYSIS FOR THE TWO-WAY CLASSIFICATION, WITH UNEQUAL NUMBERS IN THE SUBCLASSES

8. *Expectation of the mean squares for unequal numbers in each cell.*

The weighted squares of means analysis for an experiment of $i = 1, \ldots, n_i \leq n$ replicates with $j = 1, \ldots, m_i \leq m$ varieties in the ith replicate and p_{ij} sampling units in the jth variety in the ith replicate or in the ijth subplot is given below:

Source of variation	Degrees of freedom	Sum of squares	Mean squares
Replicates	$n - 1$	$\sum_{i=1}^{n} \frac{Y_{i..}^2}{p_{i.}} - \frac{Y_{...}^2}{p_{..}}$	MS(R)
Varieties or treatments	$m - 1$	$\sum_{j=1}^{m} \frac{Y_{.j.}^2}{p_{.j}} - \frac{Y_{...}^2}{p_{..}}$	MS(V)
Experimental error	$\sum_{i=1}^{n} m_i - m - n + 1$	$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{Y_{ij.}^2}{p_{ij}} - \sum_{i=1}^{n} \frac{Y_{i..}^2}{p_{i.}} - \sum_{j=1}^{m} \frac{Y_{.j.}^2}{p_{.j}} + \frac{Y_{...}^2}{p_{..}}$	MS(RV)
Sampling error	$p_{..} - \sum_{i=1}^{n} m_i$	$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \sum_{k=1}^{p_{ij}} y_{ijk}^2 - \frac{Y_{ij.}^2}{p_{ij}} \right\}$	MS(E)
Total	$p_{..} - 1$	$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{p_{ij}} y_{ijk}^2 - \frac{Y_{...}^2}{p_{..}}$	

In the above analysis

$$Y_{i..} = \sum_{j=1}^{m_i} \sum_{k=1}^{p_{ij}} y_{ijk}, \tag{8.1}$$

$$Y._{j}. = \sum_{i=1}^{n_j} \sum_{k=1}^{p_{ij}} y_{ijk}, \tag{8.2}$$

$$\text{grand total} = Y... = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{p_{ij}} y_{ijk}, \tag{8.3}$$

$p.. =$ total number of all units in the experiment

$$= \sum_{i=1}^{n} \sum_{j=1}^{m_i} p_{ij}, \tag{8.4}$$

$$p_{i}. = \sum_{j=1}^{m_i} p_{ij}, \tag{8.5}$$

$$p._{j} = \sum_{i=1}^{n_j} p_{ij}, \tag{8.6}$$

and y_{ijk} is the yield of a sampling unit. If it is assumed that the effects contributing to the yield of an individual unit are independent and additive, the linear equation

$$y_{ijk} = \mu + r_i + v_j + (rv)_{ij} + e_{ijk} \tag{8.7}$$

may be used to express the yield of an individual sampling unit where μ is the population mean, r_i is an effect common to the ith replicate, v_j an effect common to the jth variety, $(rv)_{ij}$ an effect common to the jth variety in the ith replicate, e_{ijk} an effect common to every unit, $i = 1, 2, \ldots, n_j \leq n$, $j = 1, 2, \ldots, m_i \leq m$, and $k = 1, 2, \ldots, p_{ij}$.

The expected value for the replicate mean square is

$$\sigma_e^2 + \frac{1}{n-1}\left\{ \sum_{i=1}^{n} \sum_{j=1}^{m_i} p_{ij}^2 \left[\frac{1}{p_{i}.} - \frac{1}{p..}\right] \sigma_{rv}^2 \right.$$

$$+ \left[p.. - \frac{1}{p..} \sum_{i=1}^{n} p_{i}.^2\right] \sigma_r^2 + \left[\sum_{i=1}^{n} \frac{\sum_{j=1}^{m_i} p_{ij}^2}{p_{i}.}\right.$$

$$\left.\left. - \frac{1}{p..} \sum_{j=1}^{m} p._{j}^2\right] \sigma_v^2 \right\} \tag{8.8}$$

For the case where $n_j = n$, $m_i = m$, $p_{ij} = p$, and $p.. = mnp$, the replicate mean square has the expectation

$$\sigma_e^2 + p\sigma_{rv}^2 + mp\sigma_r^2. \tag{8.9}$$

The expected value for the variety mean square is

$$\sigma_e^2 + \frac{1}{m-1}\left\{ \sum_{j=1}^{m} \sum_{i=1}^{n_j} p_{ij}^2 \left[\frac{1}{p._j} - \frac{1}{p..}\right] \sigma_{rv}^2 \right.$$

$$+ \left[p.. - \frac{1}{p..} \sum_{j=1}^{m} p._j^2\right] \sigma_v^2 + \left[\sum_{j=1}^{m} \frac{\sum_{i=1}^{n_j} p_{ij}^2}{p._j} \right.$$

$$\left. \left. - \frac{1}{p..} \sum_{i=1}^{n} p_{i.}^2 \right] \sigma_r^2 \right\}. \qquad (8.10)$$

For the case of equal numbers for all subclasses the expected value for the variety mean square is

$$\sigma_e^2 + p\sigma_{rv}^2 + np\sigma_v^2. \qquad (8.11)$$

The expected value for the replicate by variety mean square is

(where $\sum_{i=1}^{n} m_i = \sum_{j=1}^{m} n_j = n. = m.$)

$$\sigma_e^2 + \frac{1}{m. - m - n + 1}\left\{ \left[p.. - \sum_{i=1}^{n} \sum_{j=1}^{m_i} p_{ij}^2 \left[\frac{1}{p_{i.}} + \frac{1}{p._j} \right.\right.\right.$$

$$\left. \left. \left. - \frac{1}{p..}\right]\right] \sigma_{rv}^2 + \left[\frac{1}{p..} \sum_{j=1}^{m} p._j^2 - \sum_{i=1}^{n} \frac{\sum_{j=1}^{m_i} p_{ij}^2}{p_{i.}}\right] \sigma_v^2 \right.$$

$$\left. + \left[\frac{1}{p..} \sum_{i=1}^{n} p_{i.}^2 - \sum_{j=1}^{m} \frac{\sum_{i=1}^{n_j} p_{ij}^2}{p._j}\right] \sigma_r^2 \right\}. \qquad (8.12)$$

For the case of equal numbers in all categories, the replicate by variety mean square has the expectation

$$\sigma_e^2 + p\sigma_{rv}^2. \qquad (8.13)$$

The expected value for the within-subclass mean square is σ_e^2.

9. *Expectation of the mean squares for one or zero individuals in each cell.*

For the more simple case where

$$y_{ij} = \mu + r_i + v_j + (rv)_{ij}, \qquad (9.1)$$

the expected value for the replicate mean square is

$$\frac{1}{n-1}\ E\left[\sum_{i=1}^{n}\frac{Y_{i\cdot}^{2}}{m_i}-\frac{Y_{\cdot\cdot}^{2}}{m_\cdot}\right]=\sigma_{rv}^{2}+\frac{1}{n-1}\left\{\left[m_\cdot-\frac{1}{m_\cdot}\sum_{i=1}^{n}m_i^{2}\right]\sigma_r^{2}+\left[n-\frac{1}{m_\cdot}\sum_{j=1}^{m}n_j^{2}\right]\sigma_v^{2}\right\},\qquad(9.2)$$

which is equal to

$$\sigma_{rv}^{2}+m\sigma_r^{2}\qquad(9.3)$$

for all m_i equal, all n_j equal and $m_\cdot = n_\cdot = mn$.

Likewise the expected mean square for varieties is

$$\frac{1}{m-1}\ E\left[\sum_{j=1}^{m}\frac{Y_{\cdot j}^{2}}{n_j}-\frac{Y_{\cdot\cdot}^{2}}{n_\cdot}\right]=\sigma_{rv}^{2}+\frac{1}{m-1}\left\{\left[n_\cdot-\frac{1}{n_\cdot}\sum_{j=1}^{m}n_j^{2}\right]\sigma_v^{2}+\left[m-\frac{1}{n_\cdot}\sum_{i=1}^{n}m_i^{2}\right]\sigma_r^{2}\right\}\qquad(9.4)$$

which for the orthogonal case ($m_\cdot = n_\cdot = mn$, $n_j = n$, $m_i = m$) is equal to

$$\sigma_{rv}^{2}+n\sigma_v^{2}\ .\qquad(9.5)$$

The error mean square has the expectation

$$\frac{1}{m_\cdot-m-n+1}\ E\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}y_{ij}^{2}-\sum_{i=1}^{n}\frac{Y_{i\cdot}^{2}}{m_i}-\sum_{j=1}^{m}\frac{Y_{\cdot j}^{2}}{n_j}+\frac{Y_{\cdot\cdot}^{2}}{m_\cdot}\right]=\sigma_{rv}^{2}+\frac{1}{m_\cdot-m-n+1}\left\{\left[\frac{1}{n_\cdot}\sum_{i=1}^{n}m_i^{2}-m\right]\sigma_r^{2}+\left[\frac{1}{m_\cdot}\sum_{j=1}^{m}n_j^{2}-n\right]\sigma_v^{2}\right\},\qquad(9.6)$$

which for the orthogonal case is equal to σ_{rv}^{2}.

Since $MS(R)$, $MS(V)$ and $MS(RV)$ each contain estimates of the variance σ_{rv}^{2}, σ_r^{2} and σ_v^{2}, it is necessary to use all three mean squares to obtain estimates for the various variance components.

Although the above method provides unbiased estimates for the components of variance, nothing is known about the efficiency of these estimates as compared to an efficient

method, for example, the method of maximum likelihood. If the numbers in the subclasses are greatly different and if the degrees of freedom for the various mean squares are small, then the efficiency of variance components from the weighted squares of means analysis is probably low.

In replicated yield tests on hybrid corn the number of varieties is usually large and the number of replicates per variety does not vary greatly. Therefore, the unbiased estimates of the variance components from the weighted squares of means analysis are probably suitable for most purposes. This is especially true for the replicate by variety variance component and for the variety variance component. The replicate variance component, which is of least interest, is estimated the least efficiently, but even so it may be fairly efficient since the number of varieties per replicate will usually be nearly equal.

OTHER METHODS OF ANALYSIS FOR THE TWO-WAY CLASSIFICATION WITH ONE OR ZERO INDIVIDUALS IN THE SUBCLASSES

10. *The method of fitting constants.*

For the hybrid corn experiments, it is necessary to consider only the case of missing plot yields from randomized complete block experiments. The ith replicate will give yields on m_i varieties, and the jth variety total yield is obtained from the n_j replicates in which the variety was present. The yield for the jth variety in the ith replicate may be expressed by the linear equation

$$y_{ij} = \mu + r_i + v_j + (rv)_{ij}, \tag{10.1}$$

where μ is the mean yield over all plots, r_i is the effect common to the ith replicate, v_j is the effect common to the jth variety, and $(rv)_{ij}$ is an effect common to ijth plot; $i = 1, 2, \ldots, n_j \leq n$ and $j = 1, 2, \ldots, m_i \leq m$. For the randomized complete block experiment with no missing plots $n_j = n$ and $m_i = m$.

The normal equation for the mean is

$$\sum_{i=1}^{n} m_i \mu + m_1 r_1 + \ldots + m_n r_n + n_1 v_1 + \ldots + n_m v_m = Y.. \quad . \tag{10.2}$$

For the variety effects the normal equations are (for the jth variety)

$$n_j \mu + n_j v_j + r_1 + \ldots + r_{n_j} = Y._j = \sum_{i=1}^{n_j} y_{ij} \quad . \tag{10.3}$$

The normal equations for the replicate effects are (for the ith replicate)

$$m_i\mu + m_i r_i + v_1 + \ldots + v_{m_i} = Y_{i\cdot} = \sum_{j=1}^{m_i} y_{ij} \quad . \tag{10.4}$$

For this particular case, there are mn total subclasses or plots with $\sum_{i=1}^{n} (m - m_i) = \sum_{j=1}^{m} (n - n_j)$ missing yields. For purposes of simplicity and with no loss in generality let variety 1 be the variety with yields from the largest number of replicates n_1; let variety 2 be the one with yields from the next largest number of replicates n_2, etc. Thus $n \geq n_1 \geq n_2 \geq \ldots \geq n_m > 0$. Likewise let replicate 1 be the one with the largest number of varieties m_1, replicate 2 with the next largest number of varieties m_2, etc. Thus $m \geq m_1 \geq m_2 \geq \ldots \geq m_n > 0$. The two-way classification for variety yields could be represented as follows:

Varieties	Replicates						Total	
	1	2	...	i	...	n	Yield	No.
1	y_{11}	y_{21}				y_{n1}	$Y_{\cdot 1}$	n_1
2	y_{12}	y_{22}				y_{n2}	$Y_{\cdot 2}$	n_2
.	.							
.	.							
.	.							
j	y_{1j}	y_{2j}		y_{ij}		y_{nj}	$Y_{\cdot j}$	n_j
.	.							
.	.							
.	.							
m	y_{1m}	y_{2m}				y_{nm}	$Y_{\cdot m}$	n_m
Total yield	$Y_{1\cdot}$	$Y_{2\cdot}$		$Y_{i\cdot}$		$Y_{n\cdot}$	$Y_{\cdot\cdot}$	$n_{\cdot} = m_{\cdot}$
Number	m_1	m_2		m_i		m_n	$m_{\cdot}$	

In order to obtain a unique solution for the normal equations, it is necessary to impose restrictions. Two restrictions which yield a unique solution are

$$\sum_{i=1}^{n} r_i = 0 \text{ and } \sum_{j=1}^{m} v_j = 0.$$

These relations do not appear as such in the normal equations. Now, $\sum_{i=1}^{n} m_i r_i$ may be written as

$$\sum_{i=1}^{n} \left[m - (m - m_i) \right] r_i = - \sum_{i=1}^{n} (m - m_i) r_i \qquad (10.5)$$

since $\sum_{i=1}^{n} r_i = 0.$

Likewise,

$$\sum_{j=1}^{m} n_j v_j = - \sum_{j=1}^{m} (n - n_j) v_j,$$

$$\sum_{i=1}^{n_j} m_i r_i = - \sum_{i=1}^{n} (m - m_i) r_i - \sum_{i=n_j+1}^{n} m_i r_i, \qquad (10.7)$$

and

$$\sum_{j=1}^{m_i} n_j v_j = - \sum_{j=1}^{m} (n - n_j) v_j - \sum_{j=m_i+1}^{m} n_j v_j \ . \qquad (10.8)$$

Estimates of the constants in the normal equations are required. μ is obtained from the normal equation for the mean, thus

$$\mu = \frac{1}{m.} \left\{ Y.. - \sum_{i=1}^{n} r_i m_i - \sum_{j=1}^{m} n_j v_j \right\}$$

$$= \frac{1}{m.} \left\{ Y.. + \sum_{i=1}^{n} (m - m_i) r_i + \sum_{j=1}^{m} (n - n_j) v_j \right\}. \qquad (10.9)$$

Also,

$$v_j = \frac{Y._j}{n_j} - \frac{1}{m.} \left\{ Y.. + \sum_{i=1}^{n} (m - m_i) r_i + \sum_{j=1}^{m} (n - n_j) v_j \right\}$$

$$+ \frac{1}{n_j} \sum_{i=n_j+1}^{n} r_i \qquad (10.10)$$

and

$$r_i = \frac{Y_{i.}}{m_i} - \frac{1}{m.}\left\{Y.. + \sum_{i=1}^{n}(m - m_i)r_i + \sum_{j=1}^{m}(n - n_j)v_j\right\}$$

$$+ \frac{1}{m_i}\sum_{j=m_i+1}^{m} v_j . \qquad (10.11)$$

Using the regression notation of Yates (22) the reduction in the total sum of squares due to fitting the constants μ, r_i and v_j is the error or replicate by variety sum of squares. The expectation of this quantity is

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}^2 - \left\{\mu Y.. + \sum_{i=1}^{n} r_i Y_{i.} + \sum_{j=1}^{m} v_j Y_{.j}\right\}\right]$$

$$= (m. - m - n + 1)\,\sigma_{rv}^2 \qquad (10.12)$$

with $m. - m - n + 1$ degrees of freedom.

The difference in the reduction of the total sum of squares due to fitting the constants μ, r_i and v_j and the reduction due to fitting the constants μ and r_i only is the sum of squares for varieties by the method of fitting constants. Symbolically this sum of squares is

$$\mu Y.. + \sum_{i=1}^{n} r_i Y_{i.} + \sum_{j=1}^{m} v_j Y_{.j} - \sum_{i=1}^{n}\frac{Y_{i.}^2}{m_i} \qquad (10.13)$$

and has the expectation

$$(m - 1)\,\sigma_{rv}^2 + (m. - n)\,\sigma_v^2 = E\left[\sum_{j=1}^{m} v_j Y_{.j}\right] \qquad (10.14)$$

with $(m - 1)$ degrees of freedom.

Likewise, the replicate sum of squares by the method of fitting constants has the expectation

$$E\left[\mu Y.. + \sum_{i=1}^{n} r_i Y_{i.} + \sum_{j=1}^{m} v_j Y_{.j} - \sum_{j=1}^{m}\frac{Y_{.j}^2}{n_j}\right]$$

$$= (n - 1)\,\sigma_{rv}^2 + (m. - m)\,\sigma_r^2 = E\left[\sum_{i=1}^{n} r_i Y_{i.}\right] \qquad (10.15)$$

with $(n - 1)$ degrees of freedom.

The total sum of squares has the expectation

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}^2 - \frac{Y..^2}{m.}\right] + (m. - 1)\,\sigma_{rv}^2$$

$$+ \left(m. - \frac{1}{m.}\sum_{j=1}^{m} n_j^2\right)\sigma_v^2 + \left(m. - \frac{1}{m.}\sum_{i=1}^{n} m_i^2\right)\sigma_r^2 \qquad (10.16)$$

with $(m. - 1)$ degrees of freedom. As a check the error and variety sum of squares by the method of fitting constants plus the replicate sum of squares by the weighted squares of means analysis should equal the total.

11. *Missing plot formula and expectation of mean squares for* k *values missing in one group of the other classification.*

Yates (21) has described a method by which the missing or incomplete results may be calculated and the resulting tests of significance made. The missing values are computed in such a manner as to make the error sum of squares a minimum. The expectation of the variety and replicate mean squares from a randomized complete block experiment with the calculated missing plot values included in the analysis (21) will depend upon the location of the missing values in the experiment. The coefficients for the variety and replicate variance components for the respective mean squares will be $n =$ number of replicates and $m =$ number of varieties, respectively, and are the same coefficients as for the orthogonal case where no plot yields are missing or incomplete. However, the coefficient of error variance component in the variety or replicate mean square will *not* be the same as for the orthogonal case. The coefficient for the error variance component in the error mean square will be unity if the degrees of freedom are taken equal to $(n - 1)(m - 1) - k$, where k is the number of calculated yields in the experiment.

The case of $k < m$ varietal yields missing in one of the replicates is considered first. The results are applicable also to the case of $k < n$ missing yields for one variety simply by a change of symbols n and m and σ_v^2 and σ_r^2. The table of yields, y_{ij}, with the calculated missing values, x_{1t} $(t = 1, 2, \ldots, k)$, may be represented as follows for a two-way classification or for varieties and replicates in particular:

Replicates	Varieties							Totals
	1	2	...	k	k + 1	...	m	
1	x_{11}	x_{12}		x_{1k}	$y_{1,k+1}$		y_{1m}	$Y_1 . + \sum_{t=1}^{k} x_{1t}$
2	y_{21}	y_{22}		y_{2k}	$y_{2,k+1}$		y_{2m}	$Y_2 .$
.								
.								
.								
n	y_{n1}	y_{n2}		y_{nk}	$y_{n,k+1}$		y_{nm}	$Y_n .$
Totals	$Y . _1 + x_{11}$	$Y . _2 + x_{12}$	...	$Y . _k + x_{1k}$	$Y . _{k+1}$	...	$Y . _m$	$Y . . + \sum_{t=1}^{k} x_{1t}$

The analysis of the above experiment is performed in the same manner as for any m by n classification with no missing yields except that the error degrees of freedom will be k less than for the orthogonal case. The error sum of squares is

$$E = \sum_{t=1}^{k} x_{1t}^2 + \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}^2 - \frac{1}{n}\left\{\sum_{t=1}^{k} (Y . _t + x_{1t})^2 + \sum_{t=k+1}^{m} Y . _t^2\right\} - \frac{1}{m}\left\{(Y_1 . + \sum_{t=1}^{k} x_{1t})^2 + Y_2 .^2 + \ldots + Y_n .^2\right\} + \frac{1}{mn}\left\{Y . . + \sum_{t=1}^{k} x_{1t}\right\}^2 . \tag{11.1}$$

Upon minimizing this function, k linear equations in x_{11}, x_{12}, ..., x_{1k} are obtained. The linear equations are

$$x_{11}(m-1)(n-1) - (n-1)\sum_{t=2}^{k} x_{1t} = nY_1 . + mY . _1 - Y . . \tag{11.2}$$

$$x_{12}(m-1)(n-1) - (n-1)\sum_{t=1 \neq 2}^{k} x_{1t} = nY_1 . + mY . _2 - Y . . \tag{11.3}$$

$$-(n-1)\sum_{t=1}^{k-1} x_{1t} + x_{1k}(m-1)(n-1) = nY_{1.} + mY_{.k} - Y_{..} \tag{11.4}$$

Although the above equations are easily solved by iterative methods (21) the solutions for the calculated yields for the missing plots are also easy to obtain for the case of k missing values in one replicate, thus

$$x_{11} = \frac{1}{(m-k)(n-1)}\left\{nY_{1.} - Y_{..} + (m-k)Y_{.1} + \sum_{t=1}^{k} Y_{.t}\right\} \tag{11.5}$$

$$x_{12} = \frac{1}{(m-k)(n-1)}\left\{nY_{1.} - Y_{..} + (m-k)Y_{.2} + \sum_{t=1}^{k} Y_{.t}\right\} \tag{11.6}$$

.

.

.

$$x_{1k} = \frac{1}{(m-k)(n-1)}\left\{nY_{1.} - Y_{..} + (m-k)Y_{.k} + \sum_{t=1}^{k} Y_{.t}\right\}. \tag{11.7}$$

Now the calculated value x_{1t} $(t = 1, 2, \ldots, k)$ is equal to $\mu + r_i + v_j + f(rv_{ij})$, where $f(rv_{ij})$ is some function of the errors in the experiment. The calculated plot yields will be correlated to some extent with the rv_{ij} = error effect in the remainder of the experiment and with each other. The contributions of the individual errors rv_{ij} to the calculated yields for the x_{1t} and for the x_{1t} with x_{1w} $(t = 1, \ldots, k)$ are given in table 4.

The sum $n/(n-1)(m-k)$ (see table 4) multiplied by 2 gives the covariance term of the error variance component for any two calculated yields. The covariance of the calculated missing plot yield with the errors in any variety or

TABLE 4. THE CONTRIBUTION OF THE INDIVIDUAL ERRORS TO THE CALCULATED YIELDS FOR x_{1l}, x_{1k} and x_{1t} WITH $x_{1w\ (t \neq w)}$

Replicates	Varieties							Totals
	1	2	...	k	k + 1	...	m	
x_{1l}								
1	0	0		0	$1/(m-k)$		$1/(m-k)$	1
2	$1/(n-1)$	0		0	$-1/(n-1)\ (m-k)$		$-1/(n-1)\ (m-k)$	0
.								
.								
.								
n	$1/(n-1)$	0		0	$-1/(n-1)\ (m-k)$		$-1/(n-1)\ (m-k)$	0
Totals	1	0		0	0		0	1
x_{1k}								
1	0	0		0	$1/(m-k)$		$1/(m-k)$	1
2	0	0		$1/(n-1)$	$-1/(n-1)\ (m-k)$		$-1/(n-1)\ (m-k)$	0
.								
.								
.								
n	0	0		$1/(n-1)$	$-1/(n-1)\ (m-k)$		$-1/(n-1)\ (m-k)$	0
Totals	0	0		1	0		0	1
x_{1t} with x_{1w}								
1	0	0		0	$1/(m-k)^2$		$1/(m-k)^2$	$1/(m-k)$
2	0	0		0	$1/(n-1)^2\ (m-k)^2$		$1/(n-1)^2\ (m-k)^2$	$1/(n-1)^2\ (m-k)$
.								
.								
.								
n	0	0		0	$1/(n-1)^2\ (m-k)^2$		$1/(n-1)^2\ (m-k)^2$	$1/(n-1)^2\ (m-k)$
Totals	0	0	...	0	$n/(n-1)\ (m-k)^2$	...	$n/(n-1)\ (m-k)^2$	$n/(n-1)\ (m-k)$

replicate is $2(1) = 2$. The sum of squares of the coefficients for any x_{1t} is

$$\frac{m + n - k}{(n-1)(m-k)}. \tag{11.8}$$

With these results and expressing the yield of an individual plot as $y_{ij} = \mu + r_i + v_j + rv_{ij}$, it is possible to express the expectation for the mean squares in terms of the variance components σ_{rv}^2, σ_v^2 and σ_r^2.

The correction term has the expectation

$$mn\mu^2 + m\sigma_r^2 + n\sigma_v^2 + \sigma_{rv}^2 \left\{ 1 + \frac{k}{(n-1)(m-k)} \right\}. \tag{11.9}$$

The replicate sum of squares has the expectation

$$m(n-1)\sigma_r^2 + \sigma_{rv}^2 (n-1) \left\{ 1 + \frac{k}{(m-k)(n-1)} \right\} \tag{11.10}$$

with $(n-1)$ degrees of freedom.

The variety mean square has the expectation

$$n\sigma_v^2 + \sigma_{rv}^2 \left\{ 1 + \frac{k}{(n-1)(m-1)} \right\}. \tag{11.11}$$

The error sum of squares has the expectation

$$\{mn - k - m - n + 1\} \sigma_{rv}^2 \tag{11.12}$$

with $(mn - k) - m - n + 1$ degrees of freedom.

12. *Missing plot formula and expectation of mean squares for* k *missing values for different entries in each of the two-way classifications.*

In the previous section the expectations of the mean squares from a two-way classification, or a randomized complete block design in particular, were obtained for the case of k calculated yields for missing or incomplete yields which were in one group of one of the classifications. The case in which the yields for different varieties are missing in different replicates will now be considered. The table of yields with the calculated missing yields, x_{tt} $(t = 1, 2, \ldots, k)$, for a two-way classification, or for a randomized complete block in particular, may be represented as follows:

Replicates	Varieties							Totals
	1	2	...	k	k + 1	...	m	
1	x_{11}	y_{12}		y_{1k}	$y_{1,k+1}$		y_{1m}	$Y_{1\cdot} + x_{11}$
2	y_{21}	x_{22}		y_{2k}	$y_{2,k+1}$		y_{2m}	$Y_{2\cdot} + x_{22}$
.								
.								
.								
k	y_{k1}	y_{k2}		x_{kk}	$y_{k,k+1}$		y_{km}	$Y_{k\cdot} + x_{kk}$
k + 1	$y_{k+1,1}$	$y_{k+1,2}$		$y_{k+1,k}$	$y_{k+1,k+1}$		$y_{k+1,m}$	$Y_{k+1\cdot}$
.								
.								
.								
n	y_{n1}	y_{n2}		y_{nk}	$y_{n,k+1}$		y_{nm}	$Y_{n\cdot}$
Totals	$Y_{\cdot 1} + x_{11}$	$Y_{\cdot 2} + x_{22}$	...	$Y_{\cdot k} + x_{kk}$	$Y_{\cdot k+1}$	...	$Y_{\cdot m}$	$Y_{\cdot\cdot} + \sum_{t=1}^{k} x_{tt}$

The arrangement of the yields in this manner causes no loss in generality. The analysis of variance of the yields including the calculated yields is performed in the same manner as for the orthogonal randomized complete block except that k is subtracted from the error degrees of freedom (21).

Upon minimizing the error sum of squares and solving the k resulting equations, the calculated values for the missing or incomplete plot yields are

$$x_{tt} = \frac{1}{(mn - m - n)\,(mn - m - n + k)} \Big\{ (mn - m - n + k)(nY_{t\cdot} + mY_{\cdot t}) - m \sum_{t=1}^{k} Y_{\cdot t} - n \sum_{t=1}^{k} Y_{t\cdot} - (mn - m - n)Y_{\cdot\cdot} \Big\}. \tag{12.1}$$

The replicate, variety and error sums of squares have the expectations

$$(n-1)\left[m\sigma_r^2 + \sigma_{rv}^2 \left\{ 1 + \frac{k(n-1)\,(mn - m - n) + nk(k-1)}{(n-1)\,(mn - m - n + k)\,(mn - m - n)} \right\} \right], \tag{12.2}$$

$$(m-1)\left[n\sigma_v^2+\sigma_{rv}^2\left\{1+\frac{k(m-1)\ (mn-m-n)+mk(k-1)}{(m-1)\ (mn-m-n+k)\ (mn-m-n)}\right\}\right], \quad (12.3)$$

and

$$\sigma_{rv}^2\ \{mn-k-m-n+1\} \quad (12.4)$$

with $(n-1)$, $(m-1)$ and $(mn-k-m-n+1)$ degrees of freedom, respectively.

In this and the preceding section missing plot formulas were developed for two special cases and the expectations for the resulting mean squares obtained. Federer (6) has given missing plot formulas for p missing plots in each of the n replicates with no variety missing more than once, for p_1 varieties missing in one replicate and p_2 different varieties missing in another replicate, and for p_1, p_2 and p_3 different varieties missing in three replicates. No generalized missing plot formulas were presented but the general solution was indicated. From a practical standpoint it appears that any generalized formula would be too complicated for general usage and the iterative method of Yates (21) would be used instead.

13. *A simple method for obtaining the method of fitting constants analysis for a two-way classification with incomplete results.*

Although the method of fitting constants for a two-way classification with disproportionate subclass numbers is usually tedious, it may be applied to a randomized block experiment with missing plots by a simple dodge. Following the results obtained by Yates (21), the procedure for obtaining the method of fitting constants analysis for a randomized block experiment with missing plot values is:

1. Missing plot values are computed so as to make the error sum of squares a minimum [Yates (21), or sections 11 and 12].
2. With the missing plot values included in the table of yields the analysis of variance is computed in the same manner as for a randomized block experiment with no missing yields except that the error degrees of freedom are reduced by the number of calculated yields. The estimate of the error sum of squares is a minimum and will be the same as for the method of fitting constants.

3. The replicate sum of squares is computed by the method of weighted squares of means (sections 8 and 9).

$$\sum_{i=1}^{n} \frac{Y_{i.}^{2}}{m_i} - \frac{Y_{..}^{2}}{m_.} . \tag{13.1}$$

4. The total sum of squares of the observed values is obtained from the formula

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}^{2} - \frac{Y_{..}^{2}}{m_.} . \tag{13.2}$$

5. The variety sum of squares for the method of fitting constants is obtained by subtraction of the replicate (step 3) and the error (step 2) sums of squares from the total (step 4).

The analysis of variance (for a randomized complete block experiment with k missing plots) by the method of fitting constants is:

Source of Variation	Degrees of Freedom	Average Value of Mean Square
Replicates	$(n-1)$	$\sigma_{rv}^{2} + \frac{\sigma_r^{2}}{n-1}\left\{ m_. - \frac{1}{m_.} \sum_{i=1}^{n} m_i^{2} \right\} + \frac{\sigma_v^{2}}{n-1}\left\{ n - \frac{1}{m_.} \sum_{j=1}^{m} n_j^{2} \right\}$
Varieties	$(m-1)$	$\sigma_{rv}^{2} + \left[\frac{m_. - n}{m-1} \right] \sigma_v^{2}$
Error	$(m-1)(n-1) - k$	σ_{rv}^{2}

The replicate sum of squares by the method of fitting constants may be obtained in a manner similar to that just described for varieties. All these results follow from the theory given by Yates (21) and from the expectations of the mean squares given in the preceding sections.

Another procedure (21) for obtaining the method of fitting constants analysis would be the following:

1. Compute the missing plot values.
2. With the missing values inserted in the table of yields the results are orthogonal and the estimation of the regression coefficients μ, v_j and r_i proceeds in the same manner as the orthogonal case.
3. With these estimates of μ, v_j and r_i the method of analysis proceeds in the same manner as that ordinarily used for obtaining sums of squares by the method of fitting constants (see section 10).

The first method given above is quite simple and straightforward and is preferred. It was the one used for some of the experiments summarized. This method of computing the analysis of variance is designated as method of analysis number 3.

EXPERIMENTAL RESULTS AND DISCUSSION

GENERAL DESCRIPTION OF THE DATA

A number of yield trials are conducted annually in the corn improvement program of the Iowa Agricultural Experiment Station. One of the main objectives of these tests is to select the highest yielding corn strains for the various sections of Iowa. The entries in a yield trial may be single crosses, double crosses, top crosses, inbreds, open-pollinated varieties, synthetic varieties, or some other strain of corn, or they may consist of various combinations of the different types of corn strains. The field design for a yield test may be a randomized complete block, lattice, latin square, split plot or other. The yield trial may be planted at more than one place or location.

The data used in this study represent 302 corn yield trials with entries and designs varying as indicated above. These data were obtained from Dr. G. F. Sprague and Dr. J. L. Robinson of the Iowa Agricultural Experiment Station. The 232 experiments obtained from Dr. Sprague are designated as the Iowa Experimental Corn Yield Trials (IECYT). The 70 experiments obtained from Dr. Robinson are designated as the Iowa Corn Yield Tests (ICYT). The 302 experiments

were planted during the years 1940 to 1946. A description of the number of experiments representing the various categories is presented in table 5. Fifty-eight percent of the experiments were made up of double crosses. The experiments containing entries which were other than top crosses, single crosses or double crosses constitute only a small fraction (2.6 percent) of the total.

The relationship between yield and number of missing hills is approximately linear for five or less missing hills. Therefore, it is the practice at the Iowa Agricultural Experiment Station to adjust plot yields with five or less missing hills and to calculate the expected yields for the plots when there are more than five missing hills. The data from the ICYT group, and part of the experiments for 1943 and 1944 and all of the experiments for 1945 and 1946 of the IECYT were treated in this manner. In the remaining IECYT, the plot yields were adjusted for 1 to 19 missing hills. If all 20 hills were missing, missing plot formulas were used to compute the expected yields. Since the average experimental errors, for the years in which the missing plot values were calculated for plots having more than five missing hills and for the years in which the individual plot yields were adjusted by a linear relation for number of missing hills, are nearly identical [see Federer (6), tables 11 to 18], the two groups are considered as one.

TABLE 5. SOURCE AND TYPE OF DATA AND NUMBER OF EXPERIMENTS OF CORN YIELD TRIALS.

Year in which experiment was grown	Iowa experimental corn yield trials* Type of hybrid or strain chiefly concerned				Iowa corn yield tests ** Type of hybrid or strain chiefly concerned			
	Top cross	Single cross	Double cross	Other ***	Top cross	Single cross	Double cross	Other ***
	(No.)	(No.)	(No.)	(No.)	(No.)	(No.)	(No.)	(No.)
1940	20	10	14	1	0	0	8	0
1941	10	11	16	1	0	0	12	0
1942	8	9	10	0	0	0	11	0
1943	2	8	9	2	0	0	11	0
1944	0	11	14	0	0	0	8	0
1945	3	9	18	3	0	0	9	0
1946	3	15	24	1	0	0	11	0
Total	46	73	105	8	0	0	70	0

* The data for these experiments were obtained from G. F. Sprague.
** The data for these experiments were obtained from J. L. Robinson.
*** Other includes variety, synthetic, F_2 and inbreds.

EXPERIMENTAL RESULTS ON 302 CORN EXPERIMENTS FOR YIELD

14. *Analysis of the data.*

The detailed analysis for each of the 302 experiments was presented by Federer (6) in tables 11 to 18 inclusive. The randomized block analysis was used on all experiments even though the majority of the designs were lattices. Due to the fact that some of the plot yields were lost due to floods, rodents, human errors, etc., all of the experiments could not be treated in the same manner. Three types of randomized block analyses were used depending upon the number and location of missing plots. For the case of no missing plots, the analysis was the same as that for the ordinary randomized complete block; this was designated as type of analysis number 1. Type of analysis number 2 was applied to those cases for which all missing plots were in one replicate or for which the missing plots were for a different variety or strain in a different replicate (see sections 11 and 12); for this method of analysis the missing plot values were calculated and substituted in the table of yields and then the experiment was analyzed in the same manner as number 1 except that the error degrees of freedom were reduced by the number of missing plots.

Type of analysis number 3 is the method of fitting constants as described in section 13, i. e., method number 2 is followed through to obtain the error sum of squares, which is a minimum; then the replicate sum of squares is obtained by the method of weighted squares of means (section 9); and the variety sum of squares for the method of fitting constants (section 10) is obtained by subtraction of the error and the replicate sums of squares from the total sum of squares of the observed yields. The replicate mean square for type of analysis number 3 is the one obtained from the weighted squares of means analysis (see sections 8 and 9).

The calculation of the variance components is straightforward for the three methods of analysis. The replicate variance components for type of analysis number 3 were obtained from the replicate mean square calculated by the method of fitting constants. The error mean square in all three analyses is an estimate of σ_e^2. The variety mean square for methods of analysis number 1, 2 and 3 has the expectation $\sigma_e^2 + n\sigma_v^2$, $c\sigma_e^2 + n\sigma_v^2$ and $\sigma_e^2 + \frac{mn - k - n}{m - 1}\sigma_v^2$, respectively, where n equals number of replicates, m equals number of varieties, k equals number of missing plots, and c is the coefficient of σ_e^2 as described in sections 11 and 12.

The replicate mean squares are of a similar nature. The unbiased estimate of the ratio of the variety variance component, σ_v^2, to the error variance component, σ_e^2, is obtained from formulas (1.3), (1.4) and (1.5) for types of analysis number 1, number 2 and number 3, respectively.

15. *Correlations among items in the analyses of variance.*

Various correlations were obtained among the variables: number of varieties, number of replicates, error variance component, variety variance component, replicate variance component, and the mean of the experiment. The correlations among the above-listed items are given in table 6, along with the number of items used.

Considering the two groups of correlations for the IECYT and ICYT over the years 1940 through 1946, there was nothing of outstanding interest. The correlation between the error and replicate variance components is the largest. This might be expected since a more variable soil would give higher error and replicate mean squares. This certainly would be true if there were a soil gradient in one direction. The number of varieties in an experiment is related to some extent to the size of the experimental error. The number of replicates and varieties are each positively related to the size of the replicate variance component. The variety variance component is related slightly to the experimental mean, but the replicate and error variance components do not appear to have much if any relationship with the experimental mean.

Since none of the correlations from the two groups of data in table 6 are large, the majority of the variation in one of the variables is unaccounted for by the variation in the other. This is what was desired for the methods presented. High correlations among various items, for example, between number of varieties and error mean square and between error and variety variance components, would have made it necessary to modify the method of averaging ratios.

16. *Frequency distribution of variance components and of the* $\hat{a}_u$ *from 302 corn experiments.*

Frequency distributions of the error, variety and replicate variance components were obtained for the Iowa Experimental Corn Yield Trials and for the Iowa Corn Yield Tests. The patterns of the frequency distributions (see table 7) for both groups of experiments are very much alike. This might be expected since the two groups were planted in the same locality each year, and if heavy rains, say, af-

TABLE 6. CORRELATIONS BETWEEN VARIABLES OF THE CORN YIELD TRIALS.

Experimental group	Yr.	Variable	Variable					Number
			No. of varieties	No. of replicates	Error mean square	Variety variance component	Replicate variance component	
Iowa experimental corn yield trials	1940 to 1946	Error mean square	.218**	.088				232
		Variety variance component	.051	.165*	.142*			232
		Replicate variance component	.116	.200**	.419**	—.021		232
		Experiment mean			.073	.148*	.011	232
Iowa corn yield tests	1940 to 1946	Error mean square	.045	.046				70
		Variety variance component	—.154	—.210	.033			70
		Replicate variance component	.448**	.357**	.461**	—.006		70
		Experiment mean			.168	.388**	—.006	70

* Significant at 5 percent level of probability.
**Significant at 1 percent level of probability.

TABLE 7. FREQUENCY DISTRIBUTION OF THE ERROR, VARIETY AND REPLICATE VARIANCE COMPONENTS FROM 302 CORN YIELD TESTS.

Variance component											
Error				Variety				Replicate			
Class interval	ICYT	IECYT	Total	Class interval	ICYT	IECYT	Total	Class interval	ICYT	IECYT	Total
0 - 2.99	1	10	11	-3 - -0.01	1	3	4	-3 - -1.01	0	3	3
3 - 5.99	9	43	52	0 - 2.99	20	58	78	-1 - -0.01	2	20	22
6 - 8.99	15	50	65	3 - 5.99	31	57	88	0 - 0.99	19	76	95
9 - 11.99	17	43	60	6 - 8.99	13	40	53	1 - 1.99	17	49	66
12 - 14.99	9	29	38	9 - 11.99	2	22	24	2 - 2.99	8	23	31
15 - 17.99	6	22	28	12 - 14.99	0	13	13	3 - 3.99	4	14	18
18 - 20.99	3	11	14	15 - 17.99	0	13	13	4 - 5.99	4	14	18
21 - 23.99	3	5	8	18 - 20.99	1	8	9	6 - 7.99	8	13	21
24 - 26.99	0	2	2	21 - 23.99	0	4	4	8 - 9.99	2	4	6
27 - 29.99	2	2	4	24 - 26.99	1	4	5	10 - 11.99	1	3	4
30 - 32.99	2	3	5	27 - 29.99	0	1	1	12 - 13.99	1	4	5
33 - 35.99	2	3	5	30 - 32.99	0	4	4	14 - 15.99	1	2	3
36 - 38.99	1	1	2	33 - 35.99	1	0	1	16 - 17.99	0	1	1
39 - 41.99	—	0	0	36 - 38.99	—	2	2	18 - 19.99	0	1	1
42 - 44.99	—	2	2	39 - 41.99	—	1	1	20 - 21.99	0	2	2
45 - 47.99	—	0	0	42 - 44.99	—	1	1	22 - 23.99	0	1	1
48 - 50.99	—	1	1	45 - 76.99	—	0	0	24 - 25.99	0	1	1
51 - 53.99	—	1	1	77 - 77.99	—	1	1	26 - 27.99	0	0	0
54 - 62.99	—	0	0					28 - 29.99	1	0	1
63 - 65.99	—	1	1	Total	70	232	302	30 - 31.99	0	0	0
66 - 68.99	—	1	1					32 - 33.99	0	1	1
69 - 71.99	—	0	0					34 - 35.99	1	—	1
72 - 74	—	2	2					36 - 50.99	0	—	0
								51 - 51.99	1	—	1
Total	70	232	302					Total	70	232	302

fected the stands in one group they were likely to affect the stands of the other group.

The error and variety variance components give similar frequency distributions both in numerical range and in shape. The mean of the variety variance components from all experiments, 8.294, is somewhat smaller than the mean, 12.794, for the error variance component. The frequency distribution of the replicate variance components is much more closely centered around its mean, 2.955, than are the other two variance components; also, a larger proportion, 25/302, of the replicate variance components are negative.

The frequency distribution of the $\hat{a}$'s is given in table 8. Although the numerical value of the range for the unbiased estimate of the ratio of the variety to the error variance components is smaller (15.9) than either of the ranges (74 or 81) of the error or variety variance components, the range relative to the mean (16.6) is larger than either of the latter two relative ranges (5.8 and 9.8), respectively. If the variation is not related to the mean, then the variation in the $\hat{a}$'s is smaller than for the error or variety variance components.

The patterns of the distribution of the $\hat{a}$'s for top crosses, single crosses and double crosses appear to be similar. However, the pattern for top crosses appears to be a little more skewed than for the others, but this may be sampling variation resulting from the relatively small number of observations of the $\hat{a}$'s from top cross experiments. The frequency distributions (table 8) for the IECYT and the ICYT double crosses were alike in pattern and in range except for one extreme deviate in the IECYT. The extreme deviate, experiment 34, 1941 [see Federer (6), table 12], may be omitted from the comparisons since the five double crosses in this experiment were selected for wide differences in yielding ability.

The frequency distributions for all three variance components and the $\hat{a}$'s are skewed to the right. Some transformation such as the logarithmic may tend to normalize the error variance components, but such a transformation is not suitable for the variety and replicate variance components and the $\hat{a}$'s due to the negative values. A logarithmic transformation of the values $\hat{a}_u + 1$ might be used since the lower limit of $\hat{a}$ is $-1/n$.

TABLE 8. FREQUENCY DISTRIBUTION OF THE $\hat{a}$'s FROM 302 CORN EXPERIMENTS INVOLVING SINGLE CROSSES, TOP CROSSES, DOUBLE CROSSES AND OTHER KINDS OF CORN.

Class interval	Top cross	Single cross	Double cross		Other	Total
			IECYT	ICYT		
−0.4 - −0.201	0	0	1	0	0	1
−0.2 - −0.001	0	0	5	1	0	6
0.0 - 0.199	8	4	19	16	0	47
0.2 - 0.399	9	2	28	20	1	60
0.4 - 0.599	8	9	19	13	2	51
0.6 - 0.799	5	10	11	7	0	33
0.8 - 0.999	3	13	3	6	0	25
1.0 - 1.199	2	8	6	4	1	21
1.2 - 1.399	3	5	4	0	0	12
1.4 - 1.599	1	5	7	1	0	14
1.6 - 1.799	0	3	0	0	1	4
1.8 - 1.999	2	6	0	1	0	9
2.0 - 2.199	2	2	0	0	1	5
2.2 - 2.399	0	4	0	0	0	4
2.4 - 2.599	0	0	0	0	0	0
2.6 - 2.799	1	0	0	0	0	1
2.8 - 2.999	1	0	0	0	0	1
3.0 - 3.199	0	1	1	0	0	2
3.2 - 3.399	0	1	0	0	0	1
3.4 - 3.599	1	—	0	0	0	1
3.6 - 3.799	—	—	0	0	0	0
3.8 - 3.899	—	—	0	1	0	1
3.9 - 8.899	—	—	0	—	0	0
8.9 - 8.999	—	—	0	—	1	1
9.0 - 15.099	—	—	0	—	0	0
15.1 - 15.299	—	—	1	—	0	1
15.3 - 15.499	—	—	—	—	1	1
Total number	46	73	105	70	8	302

17. *Average values of the $\hat{a}$'s for single crosses, top crosses and double crosses.*

The yearly arithmetic averages of the $\hat{a}$'s for the IECYT experiments vary from 0.6999 to 1.3450, around the mean 0.9546 ± 0.1047 of the IECYT. Furthermore the mean, $\bar{a} = 0.5212 \pm 0.0648$ of the ICYT group of 70 experiments, is considerably lower than the mean of the IECYT experiments or of any of the yearly averages. This discrepancy may be explained, in part at least, by observing that the composition of the two groups of experiments is quite different, i. e., all of the entries in the ICYT experiments are double crosses while those in the IECYT group consist of several types of corn (see table 5). The mean of the $\hat{a}$'s for 46 top-cross experiments is 0.8180 ± 0.1173, for 73 single-cross experiments 1.1440 ± 0.0798, and for 105 double-cross experiments 0.6617 ± 0.1471 in the IECYT group. For all 175 double-cross experiments the mean is 0.6055. The mean ratios of the IECYT and of the ICYT double crosses differ considerably. This discrepancy may be explained by the extreme deviate obtained for experiment 34 in 1941 (6); if this experiment is omitted, then the mean ratio, $\bar{a}$, of the remaining 104 double-cross experiments in the IECYT group is 0.5224 ± 0.0478. Thus, the means of the two groups of double crosses become almost identical if the one extreme deviate is omitted. This agreement between the two groups is not unexpected since the double crosses in the ICYT have not all been selected for high yield. It was the practice of some commercial breeders to test their double-cross combinations in the ICYT. This practice has now been restricted due to the fact that only a limited number of entries may be entered and an entry may appear in only one of the four yield test districts, that is, northern, north central, south central or southern (see section 21).

A single cross, which involves two lines only, is more specific for the expression of an inherited character, such as yield, than any of the other types of corn hybrids. It is expected that the variation among single-cross yields would be the largest for any of the corn hybrids; the data confirm this. Also, the mean of the $\hat{a}$'s for the top crosses falls midway between the means for single crosses and double crosses. This result also follows expectation since the top cross may be classified as a three-way cross, as it usually is defined to be the cross of an inbred line with an unrelated tester parent. The tester may be a single cross, double cross or open-pollinated variety.

The standard errors in percent of their respective means for the various types of crosses are not excessively high. The standard errors in percent of their respective means are 14.3, 7.0, 9.2, 12.6 percent for top cross, single cross, double cross from the IECYT (omitting experiment 34, 1941), and double cross from the ICYT experiments. From a practical standpoint these means, which are unbiased estimates, may be considered accurate enough to use in determining average genetic progress.

Since the mean ratios differ, the three types of crosses will be considered separately. The average genetic progress (formula 3.1) made by selecting the highest two (tables 2 and 3) single crosses, top crosses and double crosses from a sample of m varieties for varying numbers of replicates is given in table 9. For a given number of plots, the best design for single crosses apparently is one with two replicates. On the other hand, the best design for top crosses or double crosses will depend on the number of strains in the test. From table 9 the indications are that types of corn exhibiting low genetic variation will need to be measured more accur-

TABLE 9. A COMPARISON OF POSITIVE WEIGHTED STANDARD DEVIATION UNITS FOR VARYING NUMBERS OF VARIETIES AND REPLICATES WHEN THE HIGHEST TWO YIELDING VARIETIES ARE SELECTED IN CONTRAST TO TWO RANDOMLY SELECTED ONES FROM THE SAME POPULATION (MULTIPLY EACH VALUE BY $\hat{\sigma}_e$).

Number of varieties	Average value of the two largest deviates	Positive standard deviation units with the number of replicates indicated			
		1	2	4	8
		Single-cross experiments			
25	1.74	1.359	1.552	1.686	1.767
50	2.05	1.602	1.829	1.986	2.082
75	2.22	1.734	1.981	2.151	2.254
100	2.33	1.820	2.079	2.258	2.366
150	2.48	1.938	2.213	2.403	2.518
200	2.58	2.016	2.302	2.500	2.620
		Top-cross experiments			
25	1.74	1.056	1.240	1.377	1.466
50	2.05	1.244	1.461	1.623	1.727
75	2.22	1.347	1.582	1.757	1.870
100	2.33	1.414	1.660	1.844	1.963
150	2.48	1.505	1.767	1.963	2.089
200	2.58	1.565	1.838	2.042	2.173
		Double-cross experiments			
25	1.74	0.736	0.899	1.034	1.129
50	2.05	0.867	1.059	1.218	1.330
75	2.22	0.939	1.146	1.319	1.441
100	2.33	0.986	1.203	1.384	1.512
150	2.48	1.049	1.281	1.473	1.609
200	2.58	1.092	1.332	1.533	1.674

ately than will types of corn showing high genetic variation such as single crosses.

DATA FROM EXPERIMENTS INVOLVING MORE THAN ONE TESTER

18. *Frequency distribution of yields.*

The 11 experiments involving lines, testers and replicates [see tables 10 to 12 and Federer (6), tables 23 to 27] have been discussed by Federer and Sprague (7) with regard to the relative numbers of replicates, lines and testers.

A detailed description of each of the 11 experiments is presented by Federer (6) in table 23. The analyses given by Federer and Sprague (7), table 11, were obtained using actual field weights where the plot yields had been weighed to 1/10 pound. The field weights were rounded to whole numbers in obtaining tables 10 and 12.

In a yield trial it may be desirable to determine the normality of the data. The frequency distribution of the field weights (rounded to whole numbers) for the 11 experiments was obtained (table 10). As far as was evident from observation of the histogram of yields in any of the experiments, the frequency distributions of yields do not deviate much from normal and appear to be symmetrical in form. Occasionally yield data from corn experiments are slightly skewed, but none of these 11 experiments were.

A comparison of the frequency distributions for yields and log yields indicated that the latter tended to be skewed somewhat to the left resulting in a slightly unsymmetrical distribution.

19. *Analyses of variance.*

The analyses of variance and the estimates of ratios of variance components are presented in tables 11 and 12 for yields and logarithm of yields, respectively. The error variances could not be assumed to be homogeneous, since Bartlett's (1) test of homogeneity for a set of variances gave a large chi-square value for each set. Furthermore, the variation among both the line and the line by tester variance components was larger than might be attributed to chance. The test of significance used to test the variation among variance components was an adaption of Cochran's (2) Q test, where $Q = \Sigma \frac{1}{s_i^2} (x_i - \bar{x})^2$, where a simple transformation of Q may be referred to the published tables of

TABLE 10. FREQUENCY DISTRIBUTION OF YIELDS FROM 11 EXPERIMENTS INVOLVING MORE THAN ONE TESTER.

Yield (lbs. per plot)	Randomized block design Experiment no. in								Split plot design Experiment no. in		
	1940				1941				1940		1942
	14	15	16	17	1	2	3	4	1	45	8 and 9
6		1									
7		2									
8	2	0									
9	4	1									
10	6	2									
11	6	0									
12	14	1									
13	16	1									
14	18	7							3		
15	28	2	1					1	0		
16	33	8	2	2				2	3		3
17	34	7	3	3		1		1	1		2
18	53	7	1	2		4		4	4		4
19	42	12	2	4		7		6	7		5
20	47	25	5	5		21		15	11		11
21	33	25	5	7	2	36		7	17		8
22	26	31	9	13	1	72		15	19		14
23	14	19	8	6	5	53		10	14	1	16
24	8	21	9	24	6	83		9	49	0	26
25	4	7	3	14	22	68		9	48	4	42
26	2	3	9	6	38	49	2	1	75	4	67
27		2	2	7	43	17	2	5	68	1	70
28			1	5	65	15	7	3	107	4	81
29				2	57	15	9	0	94	6	69
30					59	17	10	1	93	6	76
31					28	7	7	0	72	9	72
32					47	9	4	1	67	16	64
33					20	6	5		52	18	38
34					14	1	5		34	16	79
35					8	2	3		33	11	72
36					1	1			30	13	79
37					4	1			13	12	58
38						1			12	10	89
39									2	11	48
40									6	12	42
41									1	15	21
42									0	10	15
43									1	6	3
44										4	1
45										6	1
46										1	
47										2	
Number	390	184	60	100	420	486	54	90	936	198	1176
Exp. mean	17.83	20.42	22.47	23.35	29.14	24.50	30.52	22.28	28.98	35.93	31.92

chi-square, and where the weights are the reciprocals of the variances of the individual variance components (4). These results are not unexpected from the nature of the data, i.e., different lines and testers were compared in different locations and years, and the lines came from different kinds of parents, namely high yielding F_2 populations and open-pollinated varieties.

The unbiased estimates of the ratios of variance components, $\hat{b}$, $\hat{d}$ and $\hat{e}$, as obtained from the transformed data are slightly smaller than those obtained from the original data. The slight discrepancy would not alter the comparisons among lines, testers and replicates [see Federer and Sprague (7)] to any appreciable extent. The conclusions reached would be the same whether the data were transformed or not.

The analysis in table 11 was performed under the assumption that the yield of any individual plot of a strain of corn may be expressed in a linear manner, thus

$$y_{igh} = \mu + r_i + t_h + v_g + (vt)_{gh} + (rt)_{ih} + (rv)_{ig} + e_{igh} \quad , \quad (19.1)$$

(for a randomized complete block design; the last two terms are combined for a split-plot design) where the effects are as explained previously.

Now under the assumption that the yield effects are multiplicative rather than additive, some such model as the following might be proposed:

$$y_{igh} = \{\mu + r_i + t_h + (vt)_{gh} + (rt)_{ih} + (rv)_{ig} + e_{igh}\} v_g \quad , \quad (19.2)$$

where v_g is the line effect of the gth line or variety. To handle such a situation as this in obtaining estimates of the various mean squares the following is proposed:

$$\log y_{igh} = \log \{\mu + r_i + t_h + (vt)_{gh} + (rt)_{ih} + (rv)_{ig} + e_{igh}\} + \log v_g \quad . \quad (19.3)$$

First remove the line or variety effect as follows: the expected value of the log yield equals the observed log yield plus the experimental mean of all log yields minus the variety mean of log yields. The next step is to obtain the antilog of the expected log yields which will give the field weights of individual plot yields adjusted for line effect. The ordinary analysis of variance on yields is then obtained except that the sum of squares among line totals and the degrees of freedom for lines will be zero.

The above type of analysis was made on experiment 3 of table 11. The tester by line mean square obtained was 4.60 as compared to 4.37 in table 11. The results do not differ greatly, and this is what would be expected from the frequency distribution of yields (table 10).

TABLE 11. DATA FROM 11 CORN TOP-CROSS EXPERIMENTS INVOLVING MORE THAN ONE TESTER.

Year	Exp. No.	Kinds * of testers	Source of variation: $\hat{\sigma}_e^2$ = error variance component — Mean square	d.f.	Line x replicate — Mean square	d.f.	Line x tester — Mean square	d.f.	Line — Mean square	d.f.	No. of replicates	No. of testers	Components of variance: $\hat{\sigma}_{lt}^2$	$\hat{\sigma}_l^2$	Unbiased estimates of the ratio of variance components: $\hat{b}$	$\hat{e}$
Randomized Complete Block Designs																
1940	14	S.C.	4.13	152	4.08	152	6.08	38	65.34	38	5	2	0.390	5.931	0.091	1.418
	15	S.C.	5.11	66	9.29	66	8.84	22	25.67	22	4	2	0.932	1.581	0.169	0.304
	16	S.C.	3.97	20	3.96	20	8.76	5	42.22	5	5	2	0.958	3.347	0.197	0.769
	17	S.C.	1.50	36	3.99	36	4.32	9	23.09	9	5	2	0.564	1.628	0.344	1.031
1941	1	D.C.	3.82	138	4.60	138	8.00	69	16.41	69	3	2	1.393	1.272	0.355	0.331
	2	D.C.	5.80	160	11.08	160	7.24	80	20.57	80	3	2	0.480	1.342	0.078	0.231
	3	D.C.	2.54	16	2.04	16	4.37	8	4.19	8	3	2	0.610	0.053	0.168	0.039
	4	D.C.	4.39	28	5.78	28	6.42	14	19.99	14	3	2	0.677	2.030	0.119	0.441
Average of 8 experiments															0.190	0.570
															$\hat{b}$	$\hat{d}$
Split-plot Designs																
1940	1	S.C.	9.19	693	—	—	19.16	154	32.01	77	4	3	2.492	1.071	0.270	0.116
	45	S.C.	6.95	126	—	—	20.66	42	125.98	21	3	3	4.570	11.702	0.642	1.657
1942	8 & 9	S.C.	9.84	970	—	—	15.22	97	23.68	97	6	2	0.897	0.705	0.091	0.071
Average of 3 experiments															0.336	0.615
Average of 11 experiments															0.229	0.583

* S.C. = Single-cross testers; D.C. = double-cross testers.

TABLE 12. DATA FROM 11 CORN TOP-CROSS EXPERIMENTS INVOLVING MORE THAN ONE TESTER (YIELDS TRANSFORMED TO LOGARITHMS AND EACH MEAN SQUARE AND VARIANCE COMPONENT MULTIPLIED BY 10^4).

Year	Exp. No.	Kinds of testers	Source of variation: $\hat{\sigma}_e^2$ = error variance component		Line x replicate		Line x tester		Line		No. of replicates	No. of testers	Components of variance		Unbiased estimates of the ratio of variance components	
			Mean square	d.f.	Mean square	d.f.	Mean square	d.f.	Mean square	d.f.			$\hat{\sigma}_{lt}^2$	$\hat{\sigma}_l^2$	$\hat{b}$	$\hat{e}$
1940	14	S.C.	29	152	31	152	38	38	474	38	5	2	1.8	43.4	0.059	1.478
	15	S.C.	40	66	75	66	54	22	184	22	4	2	3.5	11.9	0.077	0.292
	16	S.C.	18	20	18	20	40	5	177	5	5	2	4.4	13.7	0.200	0.695
	17	S.C.	6	36	18	36	18	9	79	9	5	2	2.4	4.9	0.367	0.777
1941	1	D.C.	10	138	11	138	15	69	37	69	3	2	1.7	3.5	0.163	0.347
	2	D.C.	17	160	32	160	20	80	61	80	3	2	1.0	4.3	0.054	0.252
	3	D.C.	5	16	4	16	9	8	9	8	3	2	1.3	0.2	0.186	0.056
	4	D.C.	56	28	23	28	26	14	79	14	3	2	—10.0	14.3	—0.190	0.249
Average of 8 experiments															0.114	0.518
															$\hat{b}$	$\hat{d}$
1940	1	S.C.	22	693	—	—	47	154	77	77	4	3	8.3	2.5	0.375	0.113
	45	S.C.	11	126	—	—	30	42	194	21	3	3	6.3	18.2	0.558	1.628
1942	8 & 9	S.C.	22	970	—	—	34	97	50	97	6	2	6.0	1.3	0.272	0.059
Average of 3 experiments															0.402	0.600
Average of 11 experiments															0.193	0.541

Another possible model to consider would be the following:

$$y_{igh} = \{\mu + r_i + (vt)_{gh} + (rt)_{ih} + (rv)_{ig} + e_{igh}\} v_g t_h \ . \qquad (19.4)$$

In this case the line and tester effect could be removed in a manner similar to the preceding case. Although the above models[6] do not appear to fit yield data from corn hybrids, they may be satisfactory models for other types of plants or perhaps of animals.

20. *Estimation of the ratios of variance components.*

In estimating a mean effect many methods of estimation are available. The method used by Federer and Sprague (7) and in this paper has been to obtain an arithmetic mean of the unbiased estimates of a ratio. To determine the reliability of this estimate it is necessary to compare it with an efficient estimate such as the maximum likelihood estimate when it exists and is suitable for the data under consideration.

For the data of table 11 the following estimate of the mean ratio is assumed to be appropriate:

$$\bar{b}' = \frac{1}{N} \sum_{u=1}^{N} \hat{b}_u . \qquad (20.1)$$

Under the assumption that the $\hat{b}_u$ are all estimates of the same parameter β, the variance of $\bar{b}'$ is obtained from formula (2.3), thus

$$V(\bar{b}') = \frac{1}{N^2} \sum_{u=1}^{N} V(\hat{b}_u)$$

$$= \frac{2}{N^2} \sum_{u=1}^{N} \left[\frac{f_e + f_{1t} - 2}{f_{1t} f_e - 4f_{1t}}\right]_u \left[\frac{1}{n_u} + \hat{b}_u\right]^2 . \qquad (20.2)$$

If the $\hat{b}_u$ are assumed to be estimates of the β_u and the β_u are not all equal, then the variance of $\bar{b} = \frac{1}{N} \Sigma \hat{b}_u$ is

$$V(\bar{b}) = \sum_{u=1}^{N} \frac{(\hat{b}_u - \bar{b})^2}{N(N-1)} . \qquad (20.3)$$

---

[6] The reader is referred to J. W. Tukey, Biometrics 5:232, 1949, for a more recent view on the subject of non-additivity.

The ratio $V(\bar{b})/V(\bar{b}')$ is suitable as a test of significance of the difference of the ratios if f_{1t} and f_e are not too small, say less than 20 to 30. The resulting F ratio will have $N-1$ degrees of freedom associated with the numerator and infinite degrees of freedom associated with the denominator, $V(\bar{b}')$. The four examples in table 13 fit this criterion but the one in table 11 does not.

Biological theory (7) indicates that the ratios in table 11 should be different. Bartlett's (1) chi-square test of $\hat{\sigma}_e^2$'s resulted in a large value of chi-square. A chi-square test on the $\hat{\sigma}_{1t}^2$'s indicated that these variance components were unusually discrepant. Since the numerators in the ratios were different and since the denominators were different, the ratios must be different. Therefore $\hat{b}_u$ was assumed to be an estimate of β_u and all β_u are not assumed to be equal. The test suggested in the preceding paragraph does not agree with the above for the data of table 11. However, when the four experiments with $f_{1t} < 20$ are omitted (example 1, table 13), both tests agree.

Several other estimates of the mean are available. For example, if all β_u are assumed equal, a maximum likelihood estimate, $\hat{\beta}$, of β may be obtained from the equation

$$\sum_{u=1}^{N} \left\{ \frac{(nf_e)_u}{1+n_u\beta} - \frac{n_u(f_e+f_{1t})}{1+n_u\beta+\left\{\frac{f_{1t}}{f_e}\right\}_u (1+n_ub_u)} \right\} = 0. \qquad (20.4)$$

The estimate has a variance

$$V(\hat{\beta}) = \frac{N}{\sum_{u=1}^{N} \left[\frac{n_u}{1+n_u\hat{\beta}}\right]^2 \left[\frac{f_e f_{1t}}{f_e+f_{1t}+2}\right]_u}. \qquad (20.5)$$

Other estimates that might be considered are

$$\bar{b}_w = \frac{\Sigma(f_e\hat{b})_u}{\Sigma(f_e)_u}, \qquad (20.6)$$

$$\bar{b}_w' = \frac{\Sigma(f_{1t}\hat{b})_u}{\Sigma(f_{1t})_u}, \qquad (20.7)$$

and

$$\bar{b}_w^* = \frac{\Sigma(f_e+f_{1t})_u\hat{b}_u}{\Sigma(f_e+f_{1t})_u}. \qquad (20.8)$$

The variance of the above three estimates would be of the form

$$\frac{\Sigma w_i (\hat{b}_u - \bar{b}_w)^2}{N \Sigma w_i} , \qquad (20.9)$$

where w_i represents the weights appropriate for each estimate. For the data of table 11, $\bar{b}_w'$ was found to be .227, with a variance $V(\bar{b}_w') = .00225$. This variance is lower than the unweighted variance obtained for these data (see table 13).

A first approximation of an estimate of $\hat{\beta}$ in equation (20.4) may be obtained from either equation (20.6) or equation (20.8) although the latter was closer to $\hat{\beta}$ in the four examples of table 13; the values of $\bar{b}_w{}^*$ were 0.199, 0.685, 0.1046 and 2.139 for the four examples of table 13 and 0.200 for the data of table 11.

Still another type of estimate of the mean ratio of variance components is available from formula (2.4). For the data of table 11, $\bar{b}'' = \bar{\sigma}_{lt}{}^2/\bar{\sigma}_e{}^2 = 1.2694/5.2036 = 0.244$. The variance of $\bar{b}''$ may be approximated from formula (2.5). It was found to be .004146, which is about 1.6 times $V(\bar{b}) = .00257$. Similarly $\bar{d}''$ was obtained as 0.536 with a variance equal to 0.041031, which is about 1.4 times larger than $V(\bar{d}) = 0.028614$.

Now it appears that $V(\bar{b}') < V(\bar{b})$ unless the individual $\hat{b}_u$ are nearly all equal as in example 3 or if the f_{lt} are small. In the former case $V(\bar{b})$ could equal zero while $V(\bar{b}')$ could be a value quite different from zero. Likewise $V(\hat{\beta})$ lies between $V(\bar{b}')$ and $V(\bar{b})$ except in the last example, where divergent values such as $\hat{b}_u = 15.442$ cause undue effect on $V(\bar{b}')$ and $V(\bar{b})$ while they have little effect on $V(\hat{\beta})$.

It may be instructive to draw random samples from the 302 individual ratios presented by Federer (6). After drawing a large number of samples and computing the various estimates and their variances, it may be possible to generalize comparisons involving the methods of estimation. One important point observable from table 13 is that the variance may be considerably underestimated if the wrong variance formula is chosen. Further investigation of this problem would be desirable.

For the data reported here, formula (20.1) is considered to give an estimate of the mean ratio, and formula (20.3) is

TABLE 13. EXAMPLES SELECTED TO SHOW RELATIONSHIPS OF ESTIMATES OF MEAN RATIOS AND THEIR VARIANCES.

Example 1.* Experiments from table 11 for which $f_v > 20$					Example 2. First 10 experiments in the 1940 IECYT group [table 11 of thesis (6)]					Example 3. Selected uniform values of the ratios $\hat{b}$ [table 28 of thesis (6)]					Example 4. Selected divergent values of the ratios $\hat{a}$ from the 1945 IECYT group [table 16 of thesis (6)]				
Exp.** no.	n	f_e	$f_v=f_{1t}$	$\hat{a}=\hat{b}$	Exp. no.	n	f_e	f_v-f_{1v}	$\hat{a}-\hat{b}$	Exp. ** nos.	n	f_e	f_{v1}	$\hat{a}=\hat{b}$	Exp. no.	n	f_e	f_v	$\hat{a}=\hat{b}$
'40—14	5	152	38	0.091	1	4	699	233	0.514	'40—31, 32, 33	5	960	160	0.106	20	3	235	120	0.179
'40—15	4	66	22	0.169	2	3	336	168	0.306	'40—35, 36	6	1200	120	0.100	22	4	137	48	2.337
'41—1	3	138	69	0.355	3	3	336	168	0.071	'40—37, 39	5	640	80	0.106	23	3	61	35	1.875
'41—2	3	160	80	0.078	4	3	96	48	1.475	'41—9, 10	3	572	143	0.117	25	3	70	35	1.162
'40—1	4	693	154	0.270	5	3	96	48	0.460	'41—27, 28, 29	6	1800	240	0.100	9	4	239	80	2.095
'40—45	3	126	42	0.642	6	5	320	80	0.924	'42—21, 22, 23	4	432	96	0.113	14	6	175	35	0.570
'42-8&9	6	970	97	0.091	7	3	198	99	0.711	'44—22, 23	4	288	48	0.103	43	5	68	17	15.442
—	—	—	—	—	8	3	198	99	1.295	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	9	3	198	99	0.438	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	10	4	144	48	2.605	—	—	—	—	—	—	—	—	—	—
Average $= \bar{a} = \bar{b}$		$= 0.242$					$= 0.880$						.1064						3.887
Max. likelihood est. $= \hat{\alpha} = \hat{\beta}$		$= 0.195$					$= 0.608$						.1050						1.398
Aver. wt'd. by f_e		$= 0.192$					$= 0.697$						.1044						2.242
$V(\bar{a}') = V(\bar{b}')$		$= .00250$					$= .00930$						.000388						0.7877
$V(\bar{a}) = V(\bar{b})$		$= .00599$					$= .05580$						.000006						4.1323
$V(\hat{\alpha}) = V(\hat{\beta})$		$= .00341$					$= .01078$						.000862						0.0752

* For all 11 experiments $\bar{a} = 0.229$, $\hat{a} = 0.198$, $V(\bar{a}') = .00385$, $V(\bar{a}) = .00257$, and $V(\hat{a}) = .00516$. The average of the $\hat{a}_u$ weighted by f_e is 0.193.

**First number refers to year and the second to experiment number.

considered to give an estimate of the variance of the mean. However, it was deemed instructive to compare the various estimates and their variances on some experimental results. This was done in table 13.

DATA FROM EXPERIMENTS INVOLVING MORE THAN ONE PLACE OR LOCATION

21. *Description of the data.*

For purposes of conducting yield trials, Iowa is divided into four sections with three districts to a section. Districts 1, 2 and 3 make up the northern section of the state; districts 4, 5 and 6 the north central section; districts 7, 8 and 9 the south central section; and districts 10, 11 and 12 the southern section of Iowa. As a result of the performance of the entries in the Iowa Corn Yield Tests, corn hybrids or varieties are recommended for each of the four sections of Iowa; the sections are of approximately equal area.

During the years 1940 through 1946, six to nine replicates were planted for each experiment in the Iowa Corn Yield Tests [see Federer (6), table 18]. The entries were tested in all three districts of a section. Thus for a basis of comparison, three places or locations with six replicates at a place will be used as the standard in the following discussion. The design was considered to be the randomized complete block even though lattice designs were used exclusively in the Iowa Corn Yield Tests. Cochran (3) has compared the accuracy of lattice designs for corn yield trials in Iowa.

Over the period of years 1940 through 1946, 38 of the IECYT experiments [see Federer (6), table 28] were planted at more than one place or location. There were no missing plots in any of the 38 experiments at any of the locations. The analysis of these data is straightforward, and the expectation of the mean squares is the same as that given for design III of table 1. The analysis of data with missing plots at one or more of the places is not straightforward and has not been worked out. Therefore none of the ICYT experiments and several of the IECYT experiments involving more than one location could not be used for determining the relative magnitudes of the error to the variety by place and variety variance components.

The data on the 38 IECYT experiments which were planted at more than one location and had no missing yields have been given by Federer (6) in table 28. From these data the calculation of the estimated variance components, $\hat{\sigma}_{lv}^2$ and $\hat{\sigma}_{v}^2$, and of the means of the ratios $\hat{b}$ and $\hat{d}$ [formulas (1.9)

and (1.11)] was straightforward. The variance of the mean ratios $\bar{b}$ was approximated by the formula $\sum_{u=1}^{N} \frac{(\hat{b}_u - \bar{b})^2}{N(N-1)}$. The variance of $\bar{d}$ was computed from a similar formula.

22. *Relative value of combinations of strains, replicates and locations in corn yield trials.*

The means $\bar{b} = 0.2502 \pm 0.0503$ and $\bar{d} = 0.4197 \pm 0.0492$, obtained from the summarization of the data from the 38 IECYT experiments, may be expected to be applicable to various types of corn yield tests conducted under similar conditions. The results have broader implications than might first be realized. As stated in section 21, Iowa is divided into four areas with respect to corn yield tests. It might be argued that the division of the four areas is not the one yielding the most homogenous results. For example, the northern section has a shorter growing season than the remaining three sections, but within the section climatic conditions vary from west to east. It could be argued that districts 1, 2 and 3 should not be put into the same section for yield purposes. These arguments are sound *if* the experiments were to be conducted year after year. This is not the case, and the corn breeders have partially insured themselves against the year-to-year variation by selecting districts which varied with regard to climatic factors. Even though the state may be rezoned for other purposes, it is suggested that the present division of the state be retained for corn yield test purposes unless the experiments are to be repeated year after year.

The means $\bar{b}$ and $\bar{d}$ were computed from all the 38 experiments and were used to obtain tables 14 and 15. Since the composition of the 38 experiments varied with regard to kind of cross, it might be argued that the results should be treated separately for single crosses, top crosses and double crosses, for which $\bar{b} = 0.3820$, 0.4112 and 0.1323, respectively. This was not done because of the small number of experiments involved for each type of hybrid. The mean ratios $\bar{d} = 0.7138$, 0.3226 and 0.3307, respectively, for single crosses, top crosses and double crosses should be treated separately if enough observations on each type of cross were available. It would be expected that $\bar{d}$ from top-cross experiments would be intermediate to the other two values of $\bar{d}$. In any

TABLE 14. A COMPARISON OF POSITIVE WEIGHTED STANDARD DEVIATION UNITS FOR VARYING NUMBERS OF VARIETIES, LOCATIONS AND REPLICATES WHEN THE HIGHEST YIELDING VARIETY IS SELECTED IN CONTRAST TO A RANDOMLY SELECTED ONE FROM THE SAME POPULATION. (Multiply each value by $\hat{\sigma}_e$.)

Number of varieties	Average value of largest deviate	Positive standard deviation units with the number of locations or places indicated							
		1	2	3	4	5	6	7	8
					1 replicate				
25	1.97	0.640	0.809	0.904	0.966	1.010	1.043	1.069	1.089
50	2.25	0.731	0.924	1.033	1.104	1.154	1.192	1.221	1.244
75	2.40	0.780	0.985	1.101	1.177	1.231	1.271	1.302	1.327
100	2.51	0.815	1.031	1.152	1.231	1.287	1.329	1.362	1.388
150	2.65	0.861	1.088	1.216	1.300	1.359	1.403	1.438	1.465
200	2.75	0.893	1.129	1.262	1.349	1.410	1.456	1.492	1.521
300	2.90	0.942	1.191	1.331	1.422	1.487	1.536	1.574	1.604
500	3.05	0.991	1.252	1.400	1.496	1.564	1.615	1.655	1.687
					2 replicates				
25	1.97	0.764	0.927	1.010	1.061	1.095	1.120	1.139	1.154
50	2.25	0.873	1.059	1.154	1.212	1.251	1.280	1.301	1.318
75	2.40	0.931	1.130	1.231	1.293	1.334	1.365	1.388	1.406
100	2.51	0.974	1.182	1.287	1.352	1.396	1.427	1.451	1.470
150	2.65	1.028	1.248	1.359	1.427	1.473	1.507	1.532	1.552
200	2.75	1.067	1.295	1.410	1.481	1.529	1.564	1.590	1.611
300	2.90	1.125	1.365	1.487	1.562	1.612	1.649	1.677	1.699
500	3.05	1.183	1.436	1.564	1.643	1.696	1.735	1.764	1.786
					3 replicates				
25	1.97	0.825	0.980	1.055	1.099	1.129	1.150	1.166	1.178
50	2.25	0.943	1.120	1.205	1.256	1.289	1.313	1.331	1.345
75	2.40	1.006	1.194	1.285	1.339	1.375	1.401	1.420	1.435
100	2.51	1.052	1.249	1.344	1.401	1.438	1.465	1.485	1.501
150	2.65	1.110	1.319	1.419	1.479	1.518	1.547	1.568	1.584
200	2.75	1.152	1.368	1.473	1.534	1.576	1.605	1.627	1.644
300	2.90	1.215	1.443	1.553	1.618	1.662	1.693	1.716	1.734
500	3.05	1.278	1.518	1.633	1.702	1.748	1.780	1.805	1.824
					6 replicates				
25	1.97	0.904	1.043	1.106	1.142	1.166	1.182	1.194	1.204
50	2.25	1.032	1.192	1.263	1.305	1.331	1.350	1.364	1.375
75	2.40	1.101	1.271	1.348	1.392	1.420	1.440	1.455	1.466
100	2.51	1.152	1.329	1.409	1.456	1.485	1.506	1.522	1.534
150	2.65	1.216	1.403	1.488	1.537	1.568	1.590	1.606	1.619
200	2.75	1.262	1.456	1.544	1.595	1.627	1.650	1.667	1.680
300	2.90	1.331	1.536	1.628	1.682	1.716	1.740	1.758	1.772
500	3.05	1.399	1.615	1.713	1.769	1.805	1.830	1.849	1.864

event, these data illustrate the procedures for the summarization of a group of experiments.

In determining the procedure to follow in obtaining the highest yielding hybrids on the average, the formulas developed in sections 3 and 4 are applicable to experiments involving more than one location. If the highest yielding variety or hybrid is selected rather than a randomly selected one from a group of m varieties, the average genetic advance or progress or the average advance in the selection of high combining or yielding ability may be computed from formula (3.2). If one wishes to select the highest two, then take the average value of the highest two deviates (tables 2 and 3) for the value of $\bar{x}_m$.

Table 14 has been computed using formula (3.2) for average genetic advance. The comparison of the relative number of replicates and locations is the one of interest. For example, greater progress in selecting high yielding hybrids may be made by using six places and two replicates at a place rather than the standard of three places with six replicates at a place. The former scheme involves only two-thirds as many plots as the latter arrangement. Without considering the relative costs of more locations and fewer replicates, an experimenter would make better use of his efforts by using six locations and two replicates at a place rather than three locations and six replicates at a place.

The desirability of using more locations is further stressed by the fact that the yield results in one or more districts frequently are lost. For example, in district 11, which is the central district in the southern section of Iowa, no yield results were obtained in six of the last eight years [see Federer (6), table 18; no results were obtained in 1947, either]. It would be highly desirable, then, to use more districts in the southern section of Iowa since only two locations were used in the majority of years. Regardless of additional cost for planting more trials in the southern section, it appears that more places are necessary in order to obtain yield estimates for this area.

23. *Relative value of combinations of replicates and locations in corn yield tests.*

In obtaining an idea of the relative efficiency of a design with three places and six replicates at a place to various other designs involving replicates and places, table 15 was prepared. The relative efficiencies were obtained from the comparison of the error variances of a variety mean. Since the relative efficiencies represent a ratio, the error variance

of a variety mean may be expressed as

$$2\left\{\frac{\hat{\sigma}_e^2}{np}+\frac{\hat{\sigma}_{lv}^2}{p}\right\}=2\hat{\sigma}_e^2\left\{\frac{1}{np}+\frac{0.2502}{p}\right\}$$

where the symbols are as defined previously and 0.2502 equals $\overline{b}$ (see section 22). Thus by varying n and p various error variances may be obtained. The efficiencies in table 15 are given in percent of the error variance for a variety mean from three locations with six replicates at a place, which is $2\hat{\sigma}_e^2$ (0.1390). The relative efficiency of using six places with two replicates at a place to the standard is

$$\frac{\left\{\frac{1}{18}+\frac{0.2502}{3}\right\}}{\left\{\frac{1}{12}+\frac{0.2502}{6}\right\}}(100)=\left\{\frac{0.139}{0.125}\right\}(100)=111 \text{ percent.}$$

Table 15 presents several interesting features. Nine locations with one replicate at a location is as efficient as the standard three places with six replicates at a place. Furthermore, 24 replicates at one place is about one-half as efficient as the standard. To make this comparison even more striking, if an infinity of replicates were used at one place, the standard still would be the better plan to use. Since the experimenter usually would desire more than one replicate at a place, the arrangement of six locations with two replicates at a place would appear to be the preferable design. Although only two-thirds as many plots are planted and harvested, the additional cost of more locations may offset the apparent increase in efficiency. Further studies of the year effect on a variety yield and costs of planting, harvesting and travel need to be conducted in order to make concrete suggestions to be put into practice.

TABLE 15. RELATIVE EFFICIENCY (IN PERCENT) OF A RANDOMIZED COMPLETE BLOCK DESIGN AT THREE PLACES WITH SIX REPLICATES AT A PLACE TO OTHER COMBINATIONS OF REPLICATES AND PLACES.

Number of places	Number of replicates					
	1	2	3	6	12	24
1	11	19	24	33	42	48
2	22	37	48	67	83	95
3	33	56	71	100	125	143
4	44	74	95	133	167	190
5	56	93	119	167	208	238
6	67	111	143	200	250	286
7	78	130	167	233	292	333
8	89	148	191	267	333	381
9	100	167	215	300	375	429

LITERATURE CITED

1. Bartlett, M. S. Some examples of statistical methods of research in agriculture and applied biology. Jour. Roy. Stat. Soc., Suppl. 4:137-183. 1937.

2. Cochran, W. G. Problems arising in the analysis of a series of similar experiments. Jour. Roy. Stat. Soc., Suppl. 4:102-118. 1937.

3. ———— An examination of the accuracy of lattice and lattice square experiments on corn. Iowa Agr. Exp. Sta., Res. Bul. 289:400-415. 1941.

4. Crump, S. L. The estimation of variance components in analysis of variance. Biom. Bul., 2:7-11. 1946.

5. ———— The estimation of components of variance in multiple classifications. Unpublished Ph. D. thesis. Library, Iowa State College, Ames, Iowa. 1947.

6. Federer, W. T. Evaluation of variance components from a group of experiments with multiple classifications. Unpublished Ph. D. thesis. Library, Iowa State College, Ames, Iowa. 1948.

7. ———— and Sprague, G. F. A comparison of variance components in corn yield trials. I. Error, tester x line, and line components in corn top-cross experiments. Amer. Soc. Agron. Jour. 39:453-463. 1947.

8. Fisher, R. A. The design of experiments. 3rd ed. Oliver and Boyd, Ltd., London. 1942.

9. ———— and Yates, F. Statistical tables for biological, agricultural and medical research. Oliver and Boyd, Ltd., London. 1938.

10. Pearson, K. Tables for statisticians and biometricians, Part I. 2nd ed. Cambridge University Press. Cambridge, England. 1924.

11. ———— Tables for statisticians and biometricians, Part II. Cambridge University Press, Cambridge, England. 1931.

12. Perotti, J. M. Mean improvement in a normal variate under direct and indirect selection. Unpublished M. S. thesis. Library, Iowa State College, Ames, Iowa. 1943.

13. Satterthwaite, F. E. Synthesis of variance. Psychometrika, 6:309-316. 1941.

14. ———— An approximate distribution of estimates of variance components. Biom. Bul., 2:110-114. 1946.

15. Smith, H. F. The problem of comparing the results of two experiments with unequal errors. Jour. Council of Sci. and Ind. Res., 9:211-212. 1936.

16. Snedecor, G. W. Statistical Methods. 4th ed. The Collegiate Press, Inc. Ames, Iowa. 1946.

17. Sprague, G. F. and Tatum, L. A. General vs. specific combining ability in single crosses of corn. Amer. Soc. Agron. Jour., 34: 923-932. 1942.

18. Tippett, L. H. C. On the extreme individuals and the range of samples taken from a normal population. Biometrika. 17:364-387. 1925.

19. Wilks, S. S. Mathematical statistics. Princeton University Press. Princeton, N. J. 1944.

20. Winsor, C. P. and Clarke, G. L. A statistical study of variation in the catch of plankton nets. Jour. Marine Res., 3:1-34. 1940.

21. Yates, F. The analysis of replicated experiments when the field results are incomplete. Emp. Jour. Exp. Agr., 1:129-142. 1933.

22. ———— The analysis of multiple classifications with unequal numbers in the different classes. Jour. Amer. Stat. Assoc., 29: 51-66. 1934.

23. ———— Modern experimental design and its function in plant selection. Emp. Jour. Exp. Agr., 8:223-230. 1940.

24. ———— and Cochran, W. G. The analysis of groups of experiments. Jour. Agr. Sci., 28:556-580. 1938.

25. ———— and Zacopanay, I. The estimation of the efficiency of sampling, with special reference to sampling for yield in cereal experiments. Jour. Agr. Sci., 25:545-577. 1935.